

Should academics make money outside?

When, and under what circumstances, is it proper that academics should be involved with commercial enterprises? This question — always teasing — has been made more urgent by developments in the past few months which have involved academic scientists in attempts commercially to exploit the new techniques of genetic manipulation. Those concerned are likely to be by this stage keenly aware that many of their academic colleagues are deeply resentful of some of the most publicized developments. Harsh things are said over normally genteel coffee tables. There is a danger that the atmosphere of academic research will be soured, the reputations of academic institutions damaged and, perhaps worse still, the progress of research in this important field impeded. What is to be done?

The events which have made this question an important issue are memorable enough. It is now more than a year since a group of molecular biologists at the University of California, linked both by their common scientific interests and by their membership of a company called Genetech, announced that they were in a position to develop the manufacture of insulin (and some other peptide hormones) by means of genetically manipulated strains of *E. coli*. The announcement, widely acclaimed as an important development, also prompted complaints that those concerned had fallen into unacademic habits of not talking as freely to their academic colleagues as would normally have been the case. Since then, the number of prominent academic molecular biologists caught up in extramural commercial activities has increased substantially, as has the volume of somewhat embittered complaint at what is going on. The activities of the company known as Biogen, based in Switzerland and organized principally by Professor Charles Weissmann of the University of Zurich, usually excite most comment, partly because of the distinction of the academic scientists linked with it, partly because of the company's willingness to talk about itself and partly because of the somewhat breathless press conference on the prospects of manufacturing interferon with genetically altered *E. coli* held improbably at the Waldorf-Astoria Hotel in Manhattan earlier this year. Colleagues of those involved have now taken to complaining that such events are likely to undermine the integrity of academic life, to cloud the supposed single-mindedness of academic researchers in the pursuit of what might be called the truth and to create restraints on the normally free flow of information within the academic community.

Everybody will agree that the issues raised by these developments are complicated and are unlikely to be resolved quickly. Several sets of conflicting considerations interact. Perhaps the most serious impediment to a proper resolution of these difficulties is that, so far at least, too little has been done, by commercially inclined academics or by their critics, to define the issues which perplex them. No list of these could at this stage be complete, although a few bones of contention prominently stand out. It goes without saying that the issues, whatever they are, are most difficult when academics rather than, say, public employees are concerned, but this is not to say that important conflicts of interest do not arise when, for example, employees of the National Institutes of Health in the United States, or of the research councils in the United Kingdom, find themselves inclined to cash in on their expertise.

Universities are the chief testing grounds for the underlying principles because universities are proud to call themselves self-governing communities of scholars. Even if, in practice, many universities fall short of this ideal, seeming to be uneasy alliances

of warring factions (departments) or grudgingly subject to the inclinations of their administrators, it is a mark of grace that the ideal is so often held to be desirable. Does it necessarily follow that extramural commercial interests by some members of such a community are incompatible with the ideals they and their colleagues hold? The answer is complicated by the way in which, especially in recent years, many governments have urged that universities should contribute more directly to the economies which ultimately support them, and by the way in which universities themselves, for the best of educational reasons, have sought to attract teachers with practical and even continuing experience of industry. It is also important that the problem is by no means a creation of the past few years. In what sense, if any, does a molecular biologist with an interest in the success of some commercial genetic manipulation enterprise differ from, say, an academic who happens to be a successful novelist or textbook author, or from one who spends a substantial part of his time consulting for an engineering company or a drug house?

The simplest conclusion, a necessary but not sufficient condition of propriety, is that an academic should carry out the duties required of him by his contract of employment. The difficulty here is that while it is relatively easy for heads of departments to assess a person's diligence (though not his flair) in teaching, the standard requirement that he or she should also be active in research is much harder to evaluate. This, no doubt, is why many universities now require that academics should spend no more than one day a week on outside work. It is a rule of thumb which does at least ensure that an academic with outside interests has some time in which to pursue research. (In the United States, the growing zeal of those responsible for the public audit of research grant expenditure may turn out to be a tighter restriction.) Such requirements, of course, cannot ensure either the quality or the single-mindedness of a person's research, but it is a fair guess that those most able to contribute to the success of commercial enterprises outside their universities are likely also to be among the most valuable teachers and the most creative researchers. In the present wave of grumbling about the activities of molecular biologists, allegations of the neglect of simple academic duties are conspicuous by their absence.

A much more serious charge is that external commercial interests create intolerable conflicts of interest within academic laboratories. For how, if academic A is linked with some company, can academic B talk freely, over coffee or at some departmental seminar, of his own work in a related field? The difficulty is real, and those who have complained at recent developments are entirely right to say that too much of this could easily endanger not merely the social climate but the creativity of university laboratories. It is partly for this reason that many universities (but, shamefully, not all) require that academics should declare their outside interests to the heads of their departments. That is the very least that they should ask. For one thing, even though the head of a department may be apprised of his colleagues' outside interests, other researchers appear often to be unaware of what their colleagues are about. In these democratic days, is there not at least a case for asking that the disclosure of people's outside interests should be common knowledge within each closely-knit group of colleagues? And would it not be seemly and also important that heads of departments should themselves be required to spill the beans to those who work for them? That disclosure on this limited scale might be embarrassing to the academics concerned, or that



students might make trouble if they knew what their teachers were up to on the side, is unfortunate but unavoidable. Yet this, too, is merely a necessary but not a sufficient condition of seamliness.

The third and perhaps the most intractable cause of present discontent is, bluntly, that many academics are resentful of the money their colleagues stand to make from outside involvement. Even if, as seems probable, biotechnology turns out to be less like a gold mine than some people have been supposing, great disparities of income have arisen in the past and will no doubt arise in the years ahead. The problem is to decide what is seemly. If an academic should write a successful novel, his colleagues may envy him but they have no more right to complain than if he were to win a fortune in some lottery. If, on the other hand, his extramural rewards are dependent on some special knowledge derived from his position within the group of academics to which he belongs, his colleagues are rightly resentful. If his profit should in part derive from the use of university facilities, laboratories or computers for example, wider issues arise. In such circumstances, indeed, it is only equitable that a person's profit should be shared

with the institution to which he belongs. A neat way of doing this emerges from the constitution of the French company *Transgène* (see *Nature*, 10 July) where the University of Strasbourg and the Pasteur Institute will each hold a small percentage of the shares.

There are many, no doubt, who will argue that even these three conditions — diligence, open disclosure and profit-sharing where special knowledge is involved — are insufficient. Others will be attracted by more ambitious if more cumbersome ways of enabling academics to contribute to industrial development — companies or corporations set up by universities, for example. Universities differ sufficiently among themselves for the outcome of full-blooded debate on these issues to vary from place to place. There would be no harm in that. The danger lies in the virtual absence of debate. Academics grumble among themselves, but have done very little to ensure that their grumbles are openly discussed. While that reticence persists, the capacity of universities to contribute to industrial development will be impaired, while their reputation as self-governing communities of scholars will be tarnished.

Reviving (in the US) an old British recipe

The United States Congress seems bent on an exercise that resembles the re-invention of the wheel. Last week (see *Nature*, 17 July) it became known that the Department of Commerce is working on plans for the setting up of a number of "generic technology" centres and that both the Senate and the House of Representatives are ready to approve whatever scheme comes forward. Generic technology is the ugly name for what appears to be basic research related to some branch of industrial technology. Both the Department of Commerce and the Congress have it in mind that public funds should be used to launch a number of these centres, but that the industries most likely to benefit should provide an increasing proportion of the budgets of the centres, allowing the federal budget to be used for other purposes after the five years or so. Opinions, it is reported, are divided on the question whether universities should be given an important role in the management of the new centres. The antecedents of this proposal are as respectable as anybody could — the report of the level-headed Advisory Panel on Research and Development that contributed to last year's grand review of federal policy.

In reality, the new proposal in the United States is a neat amalgam of two kinds of institutions already tried in Britain. Is it too much to ask that Congress might profit from a study of the mixed success of the British experiments? The two kinds of enterprises concerned are the industrial research associations, originally recommended in the United Kingdom while the First World War was under way, and which have been over a switchback of ups and downs since their creation, and the ecology centres set up by Mr Anthony Wedgwood Benn (nor Mr Tony Benn) during his spell in the then Ministry of Technology in the late 1960s. The research associations resemble the American centres now proposed in that they were originally financed in partnership between central government and the group of industrial companies most likely to benefit. The more recently created technology centres were intended to be concerned with rather more specific problems than those which provided bread and butter for the research associations, but it was intended from the start that they would be based in universities. Again (for this is the way that governments think) the intention was that industry should take over financial responsibility for these centres in the course of time.

How well have these parallel devices worked? Most probably it is unfair to judge their success from successive British governments' opinions of them, but the research associations have lived through a succession of unpleasant ups and downs. When the wind was blowing in their direction, in the late 1950s and early 1960s, public policy stemmed from the conviction that

some of the ills of British industry might be cured if only there were a mechanism for making sure that basic research germane to important sectors of manufacturing industry was effectively undertaken. During this period, public funds were used on a modest but impressive scale and new research associations were from time to time created. They fell from favour for a variety of reasons. Some of them turned out to be second-rate. Some of them turned out to be more interested in the provision of market research than basic research. Others were undermined by the unwillingness of successful industrial subscribers with their own programmes of research and development to allow research associations more than a peripheral role. Central government, having agreed in the 1950s that the research associations should be a continuing charge on public funds, eventually changed its mind — and then found the process of making the research associations stand on their own feet more painful than it had imagined. The end result, however, is that some thirty substantial and apparently valuable research associations have survived from this half-century of initiative.

The British government's attempts to set up technology centres in universities have, by comparison, left hardly any trace. To be fair, the plan was hardly given a fair trial. Mr Anthony Wedgwood Benn (as he then was) swept into the Ministry of Technology in 1966 with a sheaf of plans for stimulating industrial research. To begin with, half a dozen centres of tribology (lubrication etc.) were set up, but plans to create a number of more broadly based centres were frustrated by the defeat of the Labour government in 1970. For what it is worth, some of these centres survive, making their way in the world with the help of funds from the University Grants Committee, the Department of Industry and fees for contract research. As such, they do not greatly differ from the industrially orientated centres for research and development that universities and polytechnics have set up on their own initiative.

What lessons should the United States Congress and the Department of Commerce draw from this experience? The first need is that they should know that it exists. Most probably, they should also reconcile themselves to providing support of some kind for the centres they plan to set up for rather longer than is now apparently foreseen. Industrial companies will pay up only when they know from experience that a new centre is doing a useful job. The successful research associations in Britain are those which have served industries (such as the water industry) where the potential members have little tradition of basic research. British experience throws little light on the usefulness of links with universities, but Congress might push in that direction.

British biotechnology boat comes home

The plan to set up a British biotechnology company, much talked of recently (*Nature*, 5 June) came into the open this week. A company has been set up with an initial capital of £12 million, between 40 and 50 per cent of which will be provided by the National Enterprise Board. It is hoped that it can be registered as Celltech, but the formalities had not been completed on Monday.

The managing director will be Mr Gerard Fairtlough, until two years ago joint managing director of Shell UK Limited with responsibility for the £300 million turnover of the petrochemical part of the enterprise, and who has since then been a division head at the NEB.

Mr Fairtlough said this week that the NEB had been looking at the opportunities in biotechnology since the early autumn of last year, when it commissioned a study of the opportunities, apparently from Professor J D Watson, director of the Cold Spring Harbor Laboratory in the United States. At that stage the NEB had not known that Sir Alf Spinks was working on a report on biotechnology for the Advisory Council on Applied Research and Development (ACARD).

The new company will be jointly owned by the NEB (the largest shareholder) and four commercial partners — the Prudential Assurance Company Limited, the Midland Bank Limited, Finance for Industry and British and Commonwealth Shipping Limited. It is intended that roughly a third of the board of the company should be molecular biologists.

The new company plans to recruit between 50 and 75 people over the next two years. It will not build a laboratory of its own but will use existing premises. The location has not been finally decided, but may well be in Cambridge.

The company has taken trouble to come to good terms with the Medical Research Council, and claims that it has a "framework for collaboration" that will allow it to place contracts with MRC laboratories and to have access to MRC "ideas". The embryo company has also been talking to two other institutions, not openly identified but probably including the Imperial Cancer Research Fund.

Mr Fairtlough says that particular trouble has been taken to study the ways in which molecular biologists might be involved with the new company. Appreciating that some may wish to make money out of their skill, and that others are offended by the notion, it is planned that the company will make commercial arrangements only with the institutions to which people belong, leaving them to decide how individuals should be paid.

At the outset, the new company will make and sell monoclonal antibodies, but will also explore longer-term projects. It is laid down that the company should make

money out of something within two or three years. The phrase "lean and hungry" is much used. The first opportunities will, it is said, be in the field of diagnostic tests.

At the same time, the NEB is anxious that the company should be financially strong enough to tackle "big things". The company will not always seek to market its own products to their ultimate users, but will be content with selling licences for manufacture by others where that seems the best course to follow.

The board of the NEB itself is said to be pleased at having launched its first "constructive project" — since the previous board resigned *en bloc* last November, the board has been primarily concerned with housekeeping. The fact that the NEB rather than the National Research Development Corporation has become the chief public sponsor of the new

biotechnology company is not only a sign of its skill but a mark of the disfavour in which the NRDC finds itself with academic scientists in general and those working for the Medical Research Council in particular.

On the question why the present British government may have sanctioned this modest investment, Mr Fairtlough said that the involvement of private capital had been a necessary condition, given "the way our guidelines have been evolving". He said that the new company would succeed if it could show that it had a "distinctive competence" and if it combined a hard-headed appraisal of the market with a capacity to make better use of people's talents. He considers that Celltech will be able to establish better relations with academic scientists than companies such as Genetech and Biogen.

Supreme Court stirs up more fear

Washington

Recombinant DNA research is moving from the frying pan to the fire. Last month's decision by the US Supreme Court to permit the patenting of live micro-organisms has started a new round of public debate about the conditions under which recombinant DNA research should be carried out.

When the National Institutes of Health formulated safety guidelines in 1976, the main concern was the potential health hazard. Attention is now concentrated on the rights of ownership and control over natural processes.

The Supreme Court ruling has crystallized various areas of concern. Several scientists, for example, have expressed their fears that the decision to allow a patent on a strain of *Pseudomonas* bacterium, developed by research workers with the General Electric Company by inserting foreign plasmids, could create barriers to the free flow of research ideas. Others disagree, claiming that the requirement to publish details of a patent ensures that the information will be publicly available. And the American Society for Microbiology is sponsoring a meeting in Washington this week to explore such areas of disagreement.

The ASM meeting will debate "the proper role of the patent laws in genetic research and development". Questions to be discussed include that of whether the law should be modified to preclude exclusivity or monopoly over a microorganism; and, more generally, whether the decision is consistent with the objectives of the microbiological and biomedical scientific communities.

Professor Emilio Schaechter, chairman of the committee organizing the ASM meeting, highlights one particular area needing attention, that of how research

using recombinant DNA techniques should be regulated in the private sector. Although NIH operate a system of voluntary compliance with their safety guidelines, they have no desire to act as a full-blown regulatory agency. NIH are considering delegating most safety decisions to local biohazard committees, and there is concern that the regulatory apparatus may be inadequate.

The Department of Labor's Occupational Safety and Health Administration is already taking steps to define its own role, including setting up a committee to outline voluntary control of the exposure of workers to processes involving recombinant DNA techniques.

Both the closeness of the Supreme Court decision — a five to four majority — and the justices' opinions on both sides provide an open invitation to Congress to look more closely at the patenting question. The four minority justices wrote "it is the role of Congress, not of the court, to broaden or narrow the reach of the patent laws". For the majority, Chief Justice Warren Burger admitted that the court's decision had been reached on a narrow interpretation of present patent law and added that Congress was free to amend the law to exclude from patent protection organisms produced by genetic engineering, or to "craft a statute specifically designed for such living things".

Pope John Paul II warned in an address to Unesco that genetic manipulation might be applied to "ends contradictory to those of humanity". Four days after the Supreme Court's decision was announced, religious groups representing Protestant, Catholic and Jewish viewpoints issued a statement which said that the questions raised by genetic engineering — who should control it, who would benefit and

who might bear adverse consequences — dealt with “the fundamental nature of human life”.

Finally, the whole issue may come under review by the newly established President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioural Research. Following an approach from the National Council of Churches, the commission is exploring whether any major reviews of the ethical implications of genetic manipulation are already in progress. If not, the subject might be selected for study by the commission when it meets in September.

David Dickson

Biotechnology

European setback

Brussels

THE European Commission's programme of research in biotechnology was seriously set back last week, when a meeting of CREST failed to endorse the programme or even to agree on a compromise. Over the past four years, the commission has developed a plan for spending £28 million over the next five years (half of it to be supplied by national governments) on a number of projects.

France was the chief stumbling block at last week's meeting of CREST, the committee which advises the European Community on research and development policy. French representatives argued for an alternative programme of fellowships to help train potential genetic engineers. The meeting of CREST was the last at which the biotechnology programme will be considered — a decision on what to do now rests with the Council of Ministers of the European Community, on which France will be able to exercise its veto.

According to reliable national sources, the council will have to choose between the commission's original plan, the French proposal and a compromise between the two supported by the United Kingdom. That would entail spending a total of £18 million over only four years and the elimination of some of the research projects proposed by the commission.

The CREST decision has caused much ill-feeling among European states. West German delegates at the meeting last week supported the French position, but more flexibly. One angry Dutch observer said last week that the smaller European states would feel misled by the larger states. An Italian official said that France had opposed the commission's plan because it wished to keep all projects with “pace” to itself.

The difficulty now is that the Council of Ministers, when considering the biotechnology programme, will be advised by an *ad hoc* scientific subcommittee of COREPER — the committee of community ambassadors in Brussels, which is well used to taking political considerations into account. It could, for example, think

of trading an agreement on the biotechnology programme for an agreement on sheepmeat.

There is nevertheless a chance that a compromise proposal may be adopted, if Germany can be persuaded to agree. The compromise would involve restricting the research projects to:

- the development of new reactors involving immobilized multi-enzyme systems;
- the transfer of genes to *Escherichia coli*, *Saccharomyces cerevisiae*, and other organisms;
- gene transfer in microorganisms and plants important to agriculture;
- safety studies.

The compromise also emphasizes mobility of manpower, in line with the French proposals, by devoting 30 per cent of the budget to training of foreign nationals in national laboratories, training of industrial scientists in university laboratories, and short courses in biohazards and safety measures. However, the French proposal does not mention safety questions, which must be satisfactorily answered if public confidence in the emerging technology is to be maintained.

In Britain, the possible demise of the European initiative is seen more as a danger signal than as an unmitigated disaster. Dr Duncan Davies, Chief Scientist at the Department of Industry, and chairman of the Cabinet Official Committee on Biotechnology which has just produced a draft White Paper on biotechnology in Britain, told the House of Lords recently that the commission programme — as it would affect the UK — would amount to only 10 per cent of the probable British effort over the next five years.

At Imperial Chemical Industries Limited, which probably has the most advanced industrial biotechnical team in Britain, the loss of the European programme is seen as less sad for ICI than for “UK Limited”. Things could be done on the European scale, said Dr Bernard Langley of ICI's Policy Group last week, “which would help everyone”. Coordinating safety standards and experimentation and developing new host-vector systems were examples.

In academic circles, where the UK Spinks report on biotechnology recommended an immediate injection of 20 new posts, and cash, the loss of commission money is taken more seriously. “Unless we have more posts soon, we are in deep trouble” said one leading biotechnologist last week. “And I see no sign at all of the Department of Education and Science taking an intelligent interest in the issue.” The loss of the European programme could be the end.

The draft version of the White Paper currently in circulation also may be little consolation to Britain's biotechnologists. According to one comment, “it offers everything short of real help”.

Robert Walgate

Yugoslav academics

No liberalization

At least for academics and scientists in Yugoslavia, there seems to be little sign of the liberalization expected to follow Tito's death.

Ivan Supek, a leading Yugoslav nuclear physicist and an active participant in the Pugwash movement, has just published in Britain a defence of his view that scientific research must be based on ethical principals, and of his resolute opposition to nuclear weapons. His views, he claims, have been consistently misrepresented by the official Communist press, which treats him, at best, as some kind of Don Quixote.

Supek, who is now head of the Institute of the History and Philosophy of Science in Zagreb, was formerly in charge of the Rudjer Boskovic Institute of Theoretical Physics in Zagreb. While working there, he watched with increasing alarm the development of a nuclear institute on the outskirts of Belgrade where, he suspected, preparations were being made for the development of nuclear weapons.

His own anti-nuclear campaign of the 1960s was, he says, only one aspect of a growing discontent among intellectuals and scientists, who felt their status had been devalued by the Party in the name of the working class. The University of Zagreb's 300th anniversary in 1969 was seized on by the academics as a unique opportunity to redress the balance.

The appearance of Supek's apologia coincides with an appeal to the world academic community by seven professors from the University of Belgrade for support in their campaign against the latest revision of the Serbian law on universities.

These lecturers were closely associated with the philosophical journal *Praxis*, founded in the mid-1960s as the organ of the Croatian Society for Philosophy. Originally in the hands of Zagreb academics who took a generally social democratic line, it gradually became associated with the more liberal-marxist Belgrade philosophers. Finally, it was closed in 1975, and eight lecturers were suspended. They were not allowed to teach, lecture or publish; they did, however, draw a reduced salary and could benefit from the free health service.

Their status will however be radically changed by the revision of Serbian law on universities, passed on 5 June. Under the previous law, (introduced in 1974) lecturers could be suspended if they “damaged social interests”. Under the new law, they can be sacked after two years' suspension. Persons already suspended for more than two years will be sacked in six months' time. The seven lecturers still under suspension (one has since found another post) maintain that this law has been brought in specifically to deal with their case. They further state that it is unconstitutional, being against the

Yugoslav Federal Labour code in which the specified grounds for dismissal do not include the nebulous charge of "damaging social interests", and because it specifically excludes any right of appeal. Most seriously, they see it as a violation of the Universal Declaration of Human Rights and the Helsinki Final Act, both of which have been ratified by Yugoslavia. They conclude that such "arbitrary interference and discrimination constitutes not only an act of harassment but an act of imperilling the very conditions of existence of a scholar".

Vera Rich

Scientists' pay

UK rumpus

A decade of discontent among scientists in the British Civil Service will come to a head on 5 August, when the dispute about salaries for 1980 will go to arbitration. The most immediate issue to be settled is whether the Civil Service will accept the latest pay offer by the Civil Service Department (CSD). Underlying the dispute, in which the Institution of Professional Civil Servants has taken most

of the weight, is whether existing methods for settling Civil Service pay should apply to scientists.

The chief humiliation in this year's pay offer is that some grades of the scientific Civil Service are to be offered no increase of salary. After a period in which the general level of salary increases in the United Kingdom has been running at something like 20 per cent, it is not surprising that passions have been roused.

Thus the salaries of those at the top of the Senior Principal Scientific Officer grade will, if the CSD proposal is unchanged by the arbitration award, remain unchanged between 1979 and 1980.

Mr William McCall, general secretary of the IPCS, has said in a circular to his members that the offer of 1980 salaries is "disgraceful". The offer, if it is upheld by arbitration, may have repercussions outside the Civil Service, and among the staffs of research councils in particular.

The pay of scientists in the British Civil Service has been a contentious issue since the early 1970s, but the professionals within the service appear to be divided in how they would like to see the present arrangements changed. The ICPS bases its complaints chiefly on the use of what is

called "pay research" to determine public servants' salaries.

This procedure entails the search for groups of scientists carrying out comparable work in the private sector of British industry. Quite independently, the use of pay research for determining civil servants' pay has recently been criticized in the United Kingdom on the grounds that the comparisons are flattering to the Civil Service and in any case take no account of the value of index-linked pensions to which retired public servants are entitled.

The IPCS objects to pay research because of the element of circularity which, it says, is likely to be introduced into the comparison because the Civil Service is the largest employer of scientists on research and development in the United Kingdom. But it also says that for some categories of senior people, suitable comparisons with private industry do not exist.

Other members of the scientific Civil Service are, on the other hand, anxious that there should be a more radical revision of the arrangements for regulating both pay and career prospects for scientists in the Civil Service. One body of opinion holds that the pay of scientists in the Civil Service compares well with that of scientists in,

Long-standing pay dispute continues

The dispute between the Civil Service Department of the British government and the Institution of Professional Civil Servants, the principal union representative of government scientists, has been rumbling on for two years. Since the early 1970s, the institution has reluctantly agreed that the pay of some categories of scientists should be determined on the recommendations of the Pay Research Unit Board, which is responsible for comparing civil servants' salaries with those of comparable groups outside.

The institution's misgivings about "pay research" for fixing the salaries of scientists stems from its belief that the Civil Service, which is the "largest single employer" of scientists working in research and development, is also a major influence on outside salaries. It also holds that pay research does not provide for the practice within the Civil Service of allowing for the promotion of individuals on the grounds of merit and says that career patterns outside are, in any case, quite different from those within the service.

The notion that civil servants should be fairly treated compared with people doing comparable civil work has been a cornerstone of Civil Service salary determinations since the Priestley Commission reported in 1956. More recently, however, the Fulton Commission on the organization of the Civil Service in 1968 advocated two further principles — the barriers

between the scientific and the administrative branches of the Civil Service should be broken down and salaries in the two branches should be linked. From 1968 until recently, the salary scales of Principal Scientific Officers and of Principals (in the administrative branch) have thus been identical.

The first breach in this principle came last year, when Principal Professional and Technical Officers (including engineers) in the Civil Service were awarded a salary increase which, although three per cent above the general increase, gave them salaries which were £729 less than those of Administrative Principals.

This year, scientists within the Civil Service come off even worse. The offer from the Civil Service Department on 26 June, based on the work of the Pay Research Unit, would have increased the salaries of all but the more senior scientists in the Civil Service by between 14.0 per cent and 25.3 per cent. Senior Principal Scientific Officers, of whom there are more than 700, were however offered no increase of salary, while it was proposed that four points should be added at the bottom of the salary scale to be occupied by future entrants.

Senior Scientific Officers and Principal Scientific Officers also fare badly, with increases at the top of the pay scales concerned of 9.1 per cent and 6.2 per cent respectively. More significantly, to many civil servants, the link with the salaries of Administrative

Principals has been broken — under the terms of the new offer from the Civil Service Department, the salaries of Principal Scientific Officers would in future be nearly £2,000 less than those of Administrative Principals.

The basis for the offer apparently rests, at least where the PSO grade is concerned, on the observation by the Pay Research Unit that in many industrial companies scientists are paid less than engineers. This has been held to justify the differential between the salaries of the Principal grades for scientists and engineers.

One of the issues in dispute between the Institution of Professional Civil Servants and the Civil Service Department is that of whether pay research can apply to the salaries of the small numbers of people in the grades of Senior Principal Scientific Officer and above. The CSD is insisting that comparisons with outside industry are possible at least for the SPSO grade.

There is also a dispute about the applicability of the agreed arbitration procedure.

Maximum salaries offered for 1980

Grade	No. in service (1979)	Salary offer	Percentage increase
Senior Principal SO*	728	£15,748	Nil
Principal SO	2,458	£12,050	6.2
Senior SO	3,830	£9,500	9.1
Higher SO	4,194	£7,900	17.3
SO	2,982	£6,400	16.7
Assistant SO	3,120	£4,925	22.2

*SO = Scientific Officer.

say, academic life, but that it would be in the long-term interest of those working as professional people within the Civil Service that something should be done to remove the barriers between the professional and administrative career tracks.

Although the Fulton Commission in 1968, in the course of its attack on the generalist as senior Civil Servant, advocated more administrative responsibility for professional civil servants, many scientists in the service are bitterly disappointed that so little has been done to implement these recommendations.

US science budget

Stability ahead?

Washington

Locked once again into the annual congressional switchback over research funds, science policy administrators in Washington have eagerly endorsed plans to place key aspects of the congressional budget review process on a multi-year basis. On 21 July, the House of Representatives passed a bill submitted by its Science and Technology Committee that would require the Administration to provide two-year (or longer) estimates of major research and development expenditures as part of its annual budget request to Congress.

The objective is that any authorization committee should then be able, if it so chose, to approve research programmes for two or more years. At present, most research spending has to be authorized annually (except for the National Institutes of Health, for which new authorizing legislation is now under consideration (*Nature*, 22 May)).

Multi-year authorizations would affect only part of the budget process. Decisions on appropriations (authority to spend) are usually more significant and these would continue to be decided annually. Even so, supporters of the new legislation hope that, if a companion bill can be coaxed from the Senate, the result will be a first step towards greater stability and continuity for research funding. Some even see it as a "pilot endeavour" for multi-year funding of federal programmes in general.

The idea appeals equally to Congress, to the Administration and to leaders of the research community. In the first place, it would cut down considerably on the amount of time and effort now required of all parties.

But as Dr Frank Press, director of the Office of Science and Technology Policy (OSTP), told a recent hearing on the bill, "by its nature, most research and development has to be carried out over a lengthy time-frame".

Some hope that the reduced commitment would provide more time for more detailed examination of specific policy proposals, and that multi-year authorization would help to improve the climate for long-range planning of research, a goal

that frequently conflicts with the need to respond to fluctuating political demands. The result should be, according to the bill's supporters, that research and development expenditure is less vulnerable to short-term changes in the fiscal climate. As things are, research and development suffers because, although it is a relatively small proportion of the overall federal budget, it represents almost a quarter of that which is not predetermined by specific laws, those concerned with social security for example. Extending the timescale of federal planning for research happens also to accord with "the budgetary philosophy of this Administration", according to Mr Bowman Cutter, executive associate director of the Office of Management and Budget (OMB).

In practice, however, multi-year planning will be difficult; the Office of Technology Assessment has told the committee that simple and flexible methods will be needed for treating inflation for example. Inflation affects research and development differently from other part of the federal budget; Dr Press has told the committee that many areas of science suffer from the "natural" built-in inflation that comes about because research becomes more sophisticated and the equipment and technical support required correspondingly more costly.

The House committee has taken Press's point. The legislation now passed by the House would require OSTP to assist and advise OMB on the evaluation of the effects of inflation factors on research and development and to suggest how they should be allowed for.

Most of this is relatively uncontroversial. Less so are proposed changes in the reporting requirements which existing legislation imposes on the OSTP and the National Science Foundation (NSF) to keep Congress up to date with the direction and intentions of official research and development policy. In particular, the new bill would dispense with OSTP's annual report to Congress on the nation's research and development. Instead, the NSF would

be required to produce a broad outlook every four years with revisions every two years. OSTP officials make little secret of their scepticism about the usefulness of such reports, hitherto required by the Presidential Science and Technology Advisory Organization Act of 1976, and the House Science Committee now appears to agree.

Some members of Congress, however, feel strongly that OSTP should be making public policy statements about the general health of research and development and the directions that it should be taking. They are thus unhappy about OSTP's apparent abdication of responsibility for long-range planning which, they claim, is explicitly required by its legislative mandate. Mr George Brown, a member of the House committee, in a minority report on the new bill, says OSTP has brought upon itself the difficulties it has encountered in preparing the reports required by law. There is reason to believe that the General Accounting Office is about to issue a report that is severely critical of OSTP on the same grounds.

David Dickson

India in space

One up so far

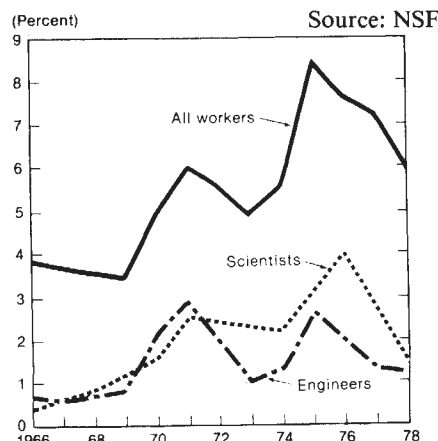
India has taken another step into the space age. Last week, the second test flight of the 17-tonne SLV-3 launcher successfully placed a small satellite into near-Earth orbit, making India one of a select group of countries with independent satellite launch facilities. SLV-3's first test flight last year ended in disaster when the rocket fell into the Indian Ocean 15 minutes after lift-off from the Sriharikota launch pad off India's east coast.

Mrs Gandhi's announcement that the second test flight was a success came during a debate in the Indian lower house (Lok Sabha) on defence ministry grants, leading some observers to speculate that the main impetus behind the rocket's development is military. India has never denied that its expertise in space technology could be used to launch reconnaissance satellites or develop intercontinental ballistic missiles, but it has always stressed that its space programme is intended to be peaceful and geared towards national development.

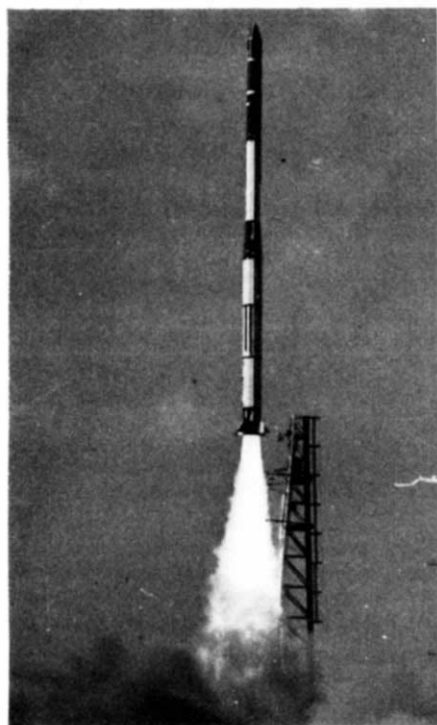
Space applications, such as remote sensing and telecommunications, form the major part of India's space programme. India has cooperated for several years with other countries such as the Soviet Union, France and the United States on remote sensing and telecommunications, one notable example being the SITE experiment with the US to bring educational television programmes to rural areas.

Rohini 1, the 40-kg satellite launched last week to record the flight characteristics of SLV-3 is the third Indian satellite to go into orbit. The Soviet Union launched the two others: Aryabhata, a scientific satellite, in 1975 and Bhaskara, for Earth observation,

US unemployment



Note: Unemployment rates are averages for the year
SOURCES: Bureau of Labor Statistics and National Science Foundation



India: where next?

in 1979. Plans for further satellites include two more in the Rohini series. These will be larger than Rohini 1 and geared to specific scientific or applications tasks. India is also planning to launch an experimental telecommunications satellite, APPLE, on the third test flight of Ariane, the European launcher, next year.

The present SLV-3 rocket, based on NASA's Scout launcher, is too small, however, to launch some of the large satellites planned. Large liquid rocket motors are already being developed at the Vikram Sarabhai Space Centre on India's west coast for incorporation into future rockets capable of putting up to 800 kg into geosynchronous orbit.

The current launcher programme, which began in 1973, grew out of earlier experience gained in building sounding rockets at the Thumba equatorial rocket launching station.

Judy Redfearn

Nuclear explosions

No alarm

Washington

When can a cloud of dust particles resemble a nuclear explosion? When it is detected by a nuclear monitoring satellite, according to a report published last week by the US Office of Science and Technology Policy.

Spurred by speculation over the nature of a mysterious explosion apparently observed by a military Vela satellite over the lower Atlantic ocean last September, OSTP set up a panel of inquiry chaired by Dr Jack Ruina of the Massachusetts Institute of Technology.

The panel was at first inclined to agree with defence officials who claimed that the satellite had detected a clandestine nuclear explosion carried out by the South African Navy either on its own initiative or in collaboration with Israel.

After intense scrutiny of the data from the satellite, however, the panel now says that the flash of light seen by the satellite was "probably not" a nuclear explosion. Its alternative hypothesis is that the satellite was hit by a small meteoroid, possibly as small as a speck of dust, and that this dislodged paint chips or other particles from the satellite's surface, which subsequently reflected sunlight into its detectors.

In support of this hypothesis, the scientists cite discrepancies in the data from the satellite's two sensors, indicating that the source may have been considerably closer to one sensor than the other.

In addition, there was a lack of any other evidence detected by stations on the ground — such as increased radiation or seismic activity — which would have corroborated the explosion hypothesis. Nor were there intelligence reports that a nuclear test had been carried out by either country. The Defense Intelligence Agency is not convinced. Before the OSTP report was released, the DIA leaked to the press its own conviction that a nuclear explosion had indeed been detected by the satellite, and that it had been carried out under weather conditions chosen to make detection difficult.

The White House, however, has adopted as its official position the OSTP panel's conclusion that there is insufficient evidence to prove a clandestine weapons test. According to State Department officials, this is the view that will prevail when a report on the incident is made to the Nuclear Non-Proliferation Treaty review meeting in Geneva next month.

David Dickson

Nuclear power

Ore running out?

The year just past was "another year of lost opportunities" according to the Nuclear Energy Agency (NEA), the OECD's own organization for the promotion of civil nuclear power. In its most recent annual report (the Eighth Activity Report covering 1979), the agency says that the growth of nuclear energy production in 1979 was a disappointing 2.2 per cent, compared with the average growth rate of 27.5 per cent a year during the period 1968-78. The pace of growth has paradoxically declined since the oil price rises of 1973.

On the face of things, however, the figures are not too bad. According to the report, there were 232 civil nuclear plants in operation at the end of 1979, and a further 229 under construction or on order, only seven more than in the previous year. New orders were almost offset by cancellation in West Germany, Spain and the United

States.

Whatever happens to new orders, the output of nuclear electricity in the OECD area in 1990 is now almost certain to be between 33 per cent and 40 per cent of the energy equivalent of oil imports in that year, according to figures quoted in the report. The hiatus in the projected growth of the nuclear industry in 1979 is attributed to the international concern with nuclear proliferation, the Three Mile Island accident, uncertainty about waste disposal and "the related problems of public understanding".

On the consequences of the Three Mile Island accident, the NEA follows a judicious line. Its report says that the accident showed that nuclear power is not accident-free, that defence in depth can work in practice but also that the accident "has revealed a number of deficiencies in the way that reactor safety has been approached up to now". It urges that more attention should be paid to the malfunctioning of other than major items of equipment, and promises to do what it can to help.

One consequence of last year's hesitation about nuclear power is, according to NEA, that the uranium mining industry has become overcautious, thus threatening future growth.

World uranium reserves at costs up to \$80/kg of uranium metal (the present economic limit on production) have increased by some 200,000 tonnes in the past two years, and now total 1.85 million tonnes. The increase came mainly from discoveries in Brazil and Canada and "improved knowledge" of deposits in Namibia, South Africa, Spain and the United States. NEA estimates that these reserves match the needs of Western nuclear power programmes between now and 2000-2005 but also considers that a significant portion of ore now identified will not be recovered "for technical or political reasons". So, the argument goes, there is an "urgent need" for the discovery of additional resources, reinforced by the wide range of the projections for nuclear power in 2025. The cumulative uranium requirements by that date range from 3.4 to 12 million tonnes, depending on nuclear power growth, the extent of reprocessing and the degree of development of fast breeders.

The NEA also looks forward to the exploitation of leaner ores, capable of yielding uranium at a cost of \$130/kg¹. Its upper estimate of the quantity of uranium in reserves now identified, not all of them explored, is put at 5.04 million tonnes, a quantity comparable with the estimated need of the nuclear industry by the second decade of the next century unless breeder technology is applied rapidly. Thus, the report concludes, the viability of the industry at the turn of the century will depend on the success of exploration in the coming ten years.

Robert Walgate

CORRESPONDENCE

Geothermal electricity

SIR, — Richardson and White¹ have raised the interesting possibility that electricity generation from geothermal aquifers in the UK might be an economic proposition. It is unfortunate, however, that they have supported their proposal with spurious comparisons; for example, comparing the performance of purpose-built (and as yet undeveloped) organic Rankine cycle generating plant with conventional unadapted district heating systems.

They state that district heating "... has dominated official thinking for several years..." and cite² Energy Paper No 9. I would not in general wish to defend that development, which is now outdated, but it actually says (p.51): "... it seems that district heating would be economic only if higher temperatures or higher housing density were to become available".

Current thinking is that the major market for heat from geothermal aquifers in the UK (as in France) will indeed be for space heating, but at the group rather than district level. Heating of only a few large buildings from each pair of wells (doublet) is envisaged, which would be located as close to the buildings as possible. Few aquifers in the UK seem likely to yield temperatures in excess of 80°C, so it will be necessary to install heating equipment capable of (1) operating at temperatures lower than those used in conventional systems and (2) extracting as much heat as possible from the geothermal fluid. These conditions have already been met in France, and suitable equipment is available. In particular, rejection temperatures can be nearer 25°C than the 60°C envisaged in ref. 1.

In the operation of a resource with a high capital cost and a low running cost resource such as geothermal energy, there is a powerful incentive to maximize the load factor. In the French installation a conventional boiler plant is used in parallel with the primary geothermal circuit, allowing geothermal to meet the "base" heating load while peak demands are met by fossil fuel. In this way, a geothermal doublet designed to meet 40% of the peak demand can typically supply 70% of the buildings' annual energy needs while achieving a load factor of 60% or more³. The relative capacities of geothermal and conventional plant will be determined by local conditions — including cost of conventional fuel at the site; analysis of the heat demand curve for similar buildings in the UK⁴ suggests that comparable figures for a UK geothermal scheme would be about 50%, 89%, and 74% respectively.

In comparing the two modes of use, we recognize that in the UK, calculations must be based on estimates and will be of value only in assessing relative rather than absolute performance. For consistency, the approach adopted by Richardson and White¹ has been used to calculate the relative performance of electricity generation and group heating for a range of probable UK aquifer temperatures and attainable flow rates. The capital cost of the doublet (£1m) and the pumping power required (250 kWe) have been assumed constant for all cases. For group heating, the geothermal system is assumed to meet 50% of the buildings' needs with the remainder supplied by conventional boilers. The size of the heat load ("equivalent dwellings" at a density in excess of 100/hectare for a mixture of commercial and residential property) is allowed to vary according to the output of the geothermal well. Capital costs for distribution mains, building internals and boiler plant are calculated according to refs 1 and 5. The rejection temperature is taken as 30°C, distribution and heat exchange losses as 10°C, the load factor as 70%, and the heat is

assumed to be sold at £2.0 per GJ. For electricity generation, the conditions given in ref. 1 have been used, while the capital cost of turbogenerator plant is assumed to be proportional to (gross output)^{1/2} with the £748,000 for a 1 MWe plant of ref. 1 used as a baseline.

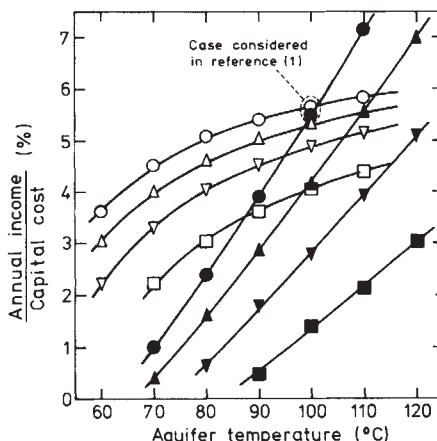


Fig.1 The effect of temperature and flow rate on the relative merits of group heating and electricity generation from geothermal aquifer sources in the UK.

Solid symbols, electricity generation; open symbols, group heating. Flow rates: ■, 10; ▽, 20; ▲, 40; ●, 50 kg s⁻¹.

The results are given in the figure. As might be expected, electricity generation becomes more attractive as temperature and flow rate increases but the return on capital is very sensitive to both parameters. Group heating is apparently less sensitive but this may be due to the method used for calculating capital costs. Comparing the data with those from French installations, capital costs derived from ref. 5 seem very high, particularly for the larger schemes. Note also that the selling price of gas, in practice, varies from £2.03 per GJ (domestic user) to £2.47 per GJ (large industrial user). The cost of useful heat, even from gas, is thus nearer £2.90 per GJ than the £1.95 per GJ quoted in ref. 1, and the cost of heat from oil would be almost double. By contrast, the selling price of electricity is fixed (at 2.2 p/kWh); this calculation might then underestimate the relative advantage of direct heat uses over power generation.

Hence the advantages of electricity generation from low-enthalpy geothermal sources are not as clear-cut as Richardson and White¹ suggest. Until a borehole is drilled at any particular site, uncertainties will exist — particularly as regards flow rate.

However, such demand for heat will occur only in the larger towns, while many of the deep aquifers in the UK underlie rural areas where group heating would not be an option. If electricity production can be shown to be an economic proposition, then the contribution of geothermal sources to UK energy supply would be greatly increased.

J.D. GARNISH

AERE Harwell
Oxon, UK

References

- Richardson, S.W. & White, A.A.L. *Nature* **286**, 103 (1980).
- Garnish, J.D. *Energy Paper No.9* (HMSO, London 1976).
- Garnish, J.D. *Pap. District Heating Association 3rd National Conf.* (1979).
- W.S. Atkins & Partners (to be published).
- Wright, J.K. et al. *Energy Paper No.20* (HMSO, London, 1977).
- Milora, S.L. & Tester, J.W. *Geothermal Energy as a Source of Electric Power*, (MIT Press, 1976).

Banking DNA sequences

SIR, — We were glad to see your article "Banking DNA Sequences" (8 May) 1980, in which you make a persuasive case that there should be such a computerized data bank. We fully agree. In fact, we already have compiled such a computerized DNA sequence bank, which now considerably exceeds the 100-kilobase figure mentioned in your article. Our DNA data bank is a natural adjunct to our Protein Sequence Reference Data Collection¹ that has been supported for the past 15 years mainly by the Institute for General Medical Sciences of the US National Institutes of Health and by the National Aeronautics and Space Administration.

Many people are aware of our facilities and we do various searches and computer studies each week for people all over the world. Our computer system includes a large number of programs especially designed for analysing such collections of data. The most frequently used capability is a search of the amino acid and nucleotide sequences for segments that match a given segment. Sequences can be retrieved quickly and compared, and statistical assessments of similarity can also be made.

Our DNA Reference Data Bank is growing continually, despite the fact that the collection of these data is aggravated by the nonavailability of material in computer-readable form (as are, for instance, the protein crystallography data). The sequences are not even in clearly readable form in some journals. We would be happy to receive sequence data in computer-readable form to add to the collection when this information is published in journals. The sequence in this additional format would then be available to others.

You also entertain the notion that it matters little by whom or where a bank is started and that a successful operation could be a depository run by a keyboard operator who sent out tapes. Based on our years of experience, we must emphatically disagree. There have been many attempts to start data collections but few have continued in operation, mainly due to greatly underestimating the amount of effort and training required for such an undertaking. Will the users be satisfied to say, in their papers, "I looked over a few sequences and reached the conclusion that..." or would they prefer to say, with confidence, "I looked over all of the sequences so far elucidated..."? Who will find corrections that have been made to the old data? Will the users want the corrected sequence in the collection? Will they want other information, such as which portions of the nucleic acid sequence code for proteins? Also, who will solicit the new material? It doesn't just drift into a depository. If the users have questions about the collection, do they want to be able to communicate with someone and expect answers? Do the users want the data center to be delicately attuned to their needs?

We feel that making the data base intellectually accessible requires a considerable amount of effort and that such an effort is warranted by the central importance of nucleic acid sequence data to many fields including medicine, genetics, and biochemistry.

M.O. DAYHOFF,
R.M. SCHWARTZ,
H.R. CHEN, L.T. HUNT,
B.C. ORCUTT, W.C. BARKER

National Biomedical Research Foundation
Georgetown University Medical Center
3900 Reservoir Road,
Washington, DC 20007

References

- Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure* Vol.5 and Suppl. 1/3 (National Biomedical Research Foundation, Washington, DC, 1972, 1973, 1976, 1979).

NEWS AND VIEWS

Just so stories and cautionary tales

from Robert M. May and Miranda Robertson

SPECULATION about the evolution of social behaviour had its historical beginning in Darwin's perplexity about (and tentative explanation for) sterile castes among the social insects. The recent upsurge of interest in such questions derives largely from Hamilton's papers in 1964 (*J. theor. Biol.* 7, 1 & 17), and their popularization in Wilson's *Sociobiology: The Modern Synthesis* (Harvard University Press, 1975). A fascinating glimpse into the sociology of science is provided by Paul Harvey's and Jon Seger's recent analysis of the citation of Hamilton's seminal papers; in the span 1964–1975, *Science Citation Index* records no more than 2–3 citations in any one year; after the appearance of Wilson's book in 1975, the citation rate explodes faster than exponentially. But Wilson's citation contains an error ("the genetical theory of social behaviour", replacing the original "the genetical evolution of social behaviour"), and this mutant form had risen to constitute 80% of all citations by 1978. Possibly as a result of the Harvey and Seger study, the mutation has not gone to fixation, and has decreased to 60% of all citations in the most recent SCI.

For a quantitative discussion of the general phenomenon of altruism, Hamilton introduced the notion of an individual's "inclusive fitness" (see *Nature* 260, 390; 1976; 260, 9; 1976). Suppose the performance of an altruistic act (such as uttering an alarm call upon noticing a predator, or, in the extreme, remaining sterile and helping raise relative's offspring) diminishes the fitness of the altruist by a factor $(1-c)$ relative to a non-altruist, but increases the fitness of the beneficiary by a factor $(1+b)$. Such altruism will be favoured by natural selection provided the cost c accruing to the performer is less than the fitness benefit b conferred upon the recipient, discounted by the degree of relatedness, r , between them;

$$r > c/b \quad (1)$$

In a diploid population with no inbreeding, sisters have $r = 1/2$, and the altruistic extreme of sterile castes ($c = 1$) will evolve only if $b > 2$. But in a simple haplo-diploid system, sisters have $r = 3/4$, and the critical degree of benefit is the weaker $b > 4/3$. This observation is one of the crucial points of Hamilton's paper.

But these ideas about inclusive fitness are based on the phenotype (altruist and non-altruist), and do not deal with the underlying genotypes and genetic mechanisms. So when ethologists, geneticists and theoreticians met recently in Berlin to talk about the evolution of social behaviour*, much discussion centred on the extent to which quick and dirty conclusions obtained from inclusive fitness arguments can be trusted. Several authors (Cavalli-Sforza & Feldman *Theor. Pop. Biol.* 14, 268; 1978; Charlesworth, Uyenoyama & Feldman, this Conference) have recently analyzed the population genetics of altruistic traits governed by one locus with two alleles. These formal, genetic studies reveal at least three shortcomings in the phenotypic inclusive fitness approach. First, the assumption that relatedness-discounted benefits and costs can be added in a simple linear fashion does not always hold; often, it will be more appropriate to compound the effects in multiplicative fashion. However, as long as both costs and benefits are relatively small, so that only weak selection operates, the difference between additive and multiplicative models is unimportant. Second, the structure of the coefficient of relatedness deduced from a genetic treatment can be different from that deduced by simple inclusive fitness arguments for the asymmetric cases of sister-brother and brother-sister in haplo-diploid species. For example, the haplo-diploid sister-brother coefficient is modified from $1/4$ to $(1/4)/(1+bh/4)$, where h measures the probability that the heterozygous genotypes Aa are altruists (the homozygotes AA being altruists, and

aa being non-altruists). Although these complications simply are not taken into account in the inclusive fitness analysis of phenotypes, the structural differences between the two approaches are again negligible in the limit of weak selection (small b and c). Third, nonlinear dynamical behaviour in the genetic models can lead to stable polymorphism, with both altruistic A and non-altruistic a alleles persisting together; in linear or additive models, be they phenotypic or genetic, if an altruistic trait can invade a population it will necessarily go to fixation.

All these differences between formally exact genetic analyses of relatedness and approximate inclusive fitness analyses are small if selection is weak. Moreover, the inclusive fitness approach provides intuitive biological understanding that is quite absent from the mathematical edifices of population genetics. Although far from unanimous, people at the conference took the view on balance that rough inclusive fitness arguments are likely to remain our main guide in thinking about the evolution of social behaviour.

More important, practical difficulties in assessing the costs c and benefits b in equation (1) will usually outweigh considerations of formal rigour. Among such difficulties identified in the conference discussions are how to combine different components (such as survival and reproductive success) into an aggregate measure of inclusive fitness, and how to determine paternity.

False assumptions about paternity, in particular, are now recognized as an important hazard in empirical tests of inclusive fitness. Clearly, the accuracy of quantitative measurements of inclusive fitness made on the basis of field observations will depend heavily on the confidence with which paternity can be assigned; and it has turned out to be a very wise ethologist indeed who knows his fathers. The important attempt by Trivers and Hare (*Science* 191, 249; 1976; see also *Nature* 260, 9; 1976) to explain the altruism of the hymenopteran worker castes by a quantitative study of data bearing on their inclusive fitness gains has been seriously

*The Dahlem Conference on the 'Evolution of Social Behaviour' was held in Berlin on February 18–22. The proceedings will be published as a book available from Dahlem Konferenzen, Delbrückstrasse 4c, D-1000, Berlin 33.

Robert M. May is the Class of 1877 Professor of Zoology at Princeton University, Miranda Robertson is an associate editor of *Nature*.

eroded by, *inter alia*, the discovery that multiple matings are common in the ant species on which they based their calculations. (Alexander & Sherman *Science*, 196, 494; 1977), and that coefficients of relationship are therefore not what they might seem.

Nor are the multiple matings of insects the only hazard to inclusive fitness analysis. Students of vertebrate behaviour have only recently begun to recognize the full extent of the gene dispersal going on behind their backs in what they fondly believed to be monogamous species. P. Marler (Rockefeller University) and M. Baker (Colorado State University) pointed out, for example, that reports of healthy clutches of young hatched by the 'monogamous' mates of vasectomized birds have been in the literature for at least five years (e.g. Bray, Kennelly & Guarino, *Wilson Bull.* 87, 187; 1975). For the reproductive strategy of the fathers of such clutches, sociobiologists have coined the term kleptogamy (or, in conversation, sneaking). These studies on bird paternity, incidentally, also testify to the complications that so often beset field experiments of this kind: for many bird species (including the redwing blackbirds studied by Bray *et al.*) infertile eggs are distinguishable by their colour from fertile ones, and it could conceivably be that the unfaithful wives chose to stray only after discovering their spouses to be infertile.

In general, field workers are becoming increasingly wary of assumptions about paternity, and wherever possible checking the degree of relationship between altruists and beneficiaries by isoenzyme analysis. This policy was adopted by R. Metcalf (University of California) in his cost-benefit analysis of the alpha and beta castes of paper wasps (*Polistes*). Female paper wasps may found nests and rear young as solitary individuals, or they may cooperate with a sister. In the case of cooperating sisters, one wasp, the alpha female, lays all or nearly all of the eggs so that the beta female is rearing nieces and nephews. However, the co-ops are so much more productive than the nests of solitaires that the inclusive fitness of the beta female is greater than that of a solitary one. Since this conclusion is based on calculations of the average number of genes a beta wasp will share with her nephews and nieces, by comparison with the average number shared with her own young, it is clearly essential to know for certain that she is the full sister of the alpha. The precaution taken by Metcalf, who analyzed the polymorphic enzymes of the wasps to check their relationships, is increasingly adopted by sociobiologists and has certainly helped to reveal unsuspected heterogeneity in presumed inbred colonies.

It has been generally assumed that the altruistic wasp, bird or beast itself does not have access to accurate information about

its relationship to its conspecifics, and simply directs its altruistic acts towards the individuals it grew up with. But strict kin selection could still work in the face of rampant polygyny, kleptogamy and general promiscuity provided that there were in fact an innate mechanism for distinguishing degrees of kinship. Greenberg's report (*Science* 206, 1095; 1979) that such a mechanism exists in the primitively eusocial bee *Lasioglossum zephyrum* was therefore generally agreed to be one of the most important recent contributions to the literature. There is overwhelming evidence that social insects recognize colony mates, but until recently no-one had managed to distinguish between innate and environmental factors in determining such recognition. Greenberg reared laboratory-bred bees of 14 different degrees of kinship in isolation and was able to show that when they were allowed to form colonies they preferentially admitted close kin to the nest, the probability of admission being roughly proportional to the degree of kinship. An exactly analogous mechanism was reported by K. E. Linsenmair (University of Würzburg) in a desert-dwelling social isopod. It is not yet known whether vertebrates have similar innate abilities. Waldman's and Adler's evidence for kin recognition in frogs (*Nature* 282, 611; 1979) does not distinguish between innate and environmental mechanisms; but if innate mechanisms turn out to be widespread, this will limit the applicability of J. Maynard Smith's (Sussex University) ingenious adaptations of game theory to define evolutionarily stable strategies (ESSs) for behaviour, one of whose underlying assumptions is usually that kin recognition is based on simple proximity. The whole question of 'hidden' assumptions in evolutionary theories of behaviour is clearly a sore point with empiricists, who expressed the view that

theoreticians should make them more explicit. For example, the majority of empiricists are probably not aware that most (though not all) of the ESS calculations in the current literature assume haploidy, and therefore strictly apply to nothing higher than some insects. Most of the conclusions can, however, be shown to apply to sexual diploids; and given the inevitability of simplifying assumptions, probably the most useful thing a theoretician can do is try to give some indication of how sensitive his theory is to violation of its assumptions.

On the other hand, it is incumbent on the empiricists to produce data good enough to provide a serious test of the theory. It is the shallowness of some of the earlier ethological investigations that has led to their being castigated by some as evolutionary Just-So stories. Lively discussion centred on several interesting but unfinished studies in which individual males seem to adopt one of two or more alternative reproductive strategies (for an example, fighting to maintain a territory and/or a harem *versus* kleptogamy). Such studies have so far left open many questions. Are the different phenotypes associated with different genotypes, or are they dictated by environmental constraints? Do the different strategies correspond to different categories of animal, or can, for example, a male start his adult life as a 'sneak' and later progress to ownership of a harem?

A fundamental question that has been raised regularly by the principal detractors of what has become known as sociobiology, and often ignored by its practitioners, is: what evidence is there that a particular phenotype is, in fact, adaptive? Feldman quoted the example of meiotic drive as a mechanism for preserving maladaptive genes in a population. Other extreme examples of this sort are known or suspected, and there is every reason to



100 years ago

NATURE WELCOMES SCIENCE

We take the following from the New York *Nation*:—"For the English-speaking race, wherever planted, we should have supposed NATURE to be a sufficient scientific medium, and entitled to universal support. We are partly confirmed in this view by the quotations from NATURE in the first number of *Science*, a quarto weekly journal, edited by Mr John Michels, and published at 229 Broadway, in this city. Nevertheless, the editor's statement that the enterprise has been begun 'after consultation with many of

the leading scientists in this country,' and his list of co-labourers seem to point to a real want, and to entitle this new 'record of scientific progress' to a friendly welcome. Its present size is sixteen pages, including the advertising sheet. The opening article, on the United States Naval Observatory, is from the pen of Prof. E.S. Holden." We wish our new contemporary every success, and trust that it may be the means of spreading a wide interest in science on the other side of the water.

A violent shock of earthquake occurred at Manila and throughout the Island of Luzon on July 18, which did immense damage, totally destroying several government buildings and other houses. Some of the native inhabitants were killed, but no Europeans suffered any injury. A slight shock was felt also on the 17th inst.

from *Nature* 22, 24 July, 276 & 276, 1880.

suppose that selectively neutral behaviour can become fixed in a population by historical or evolutionary accident.

To paraphrase one of the more vociferous participants, the Romantic Age of sociobiology, with its excitement and imaginative speculations about the explanatory power of inclusive fitness notions, is giving way to a more sober appreciation of the experimental difficulties that must be overcome before contending theories can be sorted out. For many, the meeting helped to crystallize thoughts about assembling data of sufficient resolution to distinguish between theories, and to draw up a tentative agenda of theoretical and empirical problems that

deserve attention. In this context, the generic format of the Dahlem Conferences (of which there have been some 20-odd, on a diversity of topics) deserves mention: canonically teutonic, with groups and subgroups forming and reforming in intricate, pre-ordained patterns, their most obvious function is to produce a book, whatever betide. But the groups at the sociobiology meeting were skillfully chosen, blending theorists and empiricists, optimists and pessimists, romantics and realists; most important, the four rapporteurs (Feldman, Harvey, Krebs, Oster) managed to extract from the discussions reports that seem to be useful signposts. □

represents an intermediate in the normal enzymatic reaction which is stable in the absence of a suitable substrate, or an autophosphorylation which may affect function is not clear. pp60^{src} also undergoes an apparent autophosphorylation reaction when in soluble form both *in vitro* and *in vivo*, and in addition is phosphorylated by the normal cellular cyclic AMP-dependent protein kinase at a separate site in the molecule^{14,9}. Whether these phosphorylation reactions modulate function is again not clear. There is less evidence that the protein kinase activity is due to viral protein for viruses other than avian sarcoma and there are some reports that activity may not be present in purified forms of SV40 and hybrid SV40 large T antigen^{18,23}. However, large T antigen is different in many ways from the other viral products which show kinase activity. Kinase activity is absent from immunoprecipitates when transformation-defective mutants of adenovirus, SV40 and polyoma are used, consistent with the conclusion that the transforming activity is a kinase although the possibility of an association between the viral protein and a cellular kinase remains.

That transformation-related protein kinases catalyze the phosphorylation of tyrosine residues in their substrates was first shown by Hunter and co-workers² in the polyoma system, where kinase activity is mainly associated with a tumour antigen of intermediate size (middle T). Phosphorylation occurs on the phenolic hydroxyl group of tyrosine. Protein kinases associated with other viral transforming gene products also phosphorylate tyrosine residues³⁻⁵. Previously, phosphorylated tyrosine residues were not known to occur in proteins, although modifications such as nucleotidyl derivatives of tyrosine are found in bacterial enzymes (for example, ref. 24). As each of these kinases associated with viral transforming gene products catalyzes the phosphorylation of tyrosine residues it seems probable that a unique class of protein kinases may be involved in the transformation of cells to the malignant state. This feature may be extremely useful in locating other kinases with similar or related functions, and in locating and measuring the phosphorylation of potential substrates.

By itself, the fact that phosphorylation occurs on tyrosine rather than on serine or threonine residues gives no clue to the molecular mechanism of transformation. One possibility is that the phosphorylation of cellular substrate proteins by the viral kinase (or by a cellular kinase which is activated by association with the viral protein) leads to functional alterations which initiate the transformation process. Since this is a highly pleiotropic response, many substrates could be involved. Kinase action *in vivo* is indicated by an approximately 10-fold increase in the phospho-

Malignant transformation and protein phosphorylation

from Thomas Langan

DURING the past decade protein phosphorylation reactions have been found to function in the regulation of a large number of cellular processes. In addition to extensively studied systems such as cyclic AMP-dependent phosphorylation, many other less well-characterized protein kinases and protein phosphorylation reactions occur, suggesting that regulation by protein phosphorylation is more wide-spread than currently appreciated. Recent work, stemming from the original observations of Collett and Erickson¹, has indicated that certain proteins responsible for transformation of virally infected cells to the malignant state are protein kinases, or are closely associated with protein kinases. These kinases have now been shown to have the novel property of catalyzing the phosphorylation of tyrosine residues in their substrate proteins, rather than the usual serine or threonine residues²⁻⁵. The findings indicate that a unique class of protein kinases may catalyze a key reaction in the production of certain types of malignancy, and there are hints that similar kinases may participate in normal regulation of cell proliferation.

The best studied system is that of the Rous or avian sarcoma virus, an RNA virus in which a single viral gene, *src*, has been shown to be responsible for transformation of cells to the malignant state⁶. A 60,000 MW protein present in transformed cells has been identified as the product of this gene by the use of deletion and temperature-sensitive mutants and, more convincingly, by direct translation *in vitro* from fragments of viral RNA containing the *src* gene^{7,8}. The protein

(pp60^{src}) is phosphorylated in two distinct regions of the molecule⁹ and as with certain other phosphoproteins, pp60^{src} seems to be itself a protein kinase. This activity is of the cyclic AMP-independent type and was originally revealed by the ability of pp60^{src} to catalyze transfer of phosphate from ATP to the heavy chain of IgG in immunoprecipitates of viral protein, prepared using antisera from tumour bearing animals^{1,10}. The reaction is clearly not of functional significance, but the ability to phosphorylate a non-physiological substrate is very advantageous for assay purposes. Extensive studies with deletion and temperature sensitive mutants^{1,10-12} and the demonstration of protein kinase activity in highly purified preparations of pp60^{src}^{13,14} as well as in pp60^{src} made by *in vitro* translation of viral RNA^{15,16} show that the protein kinase activity is a property of the *src* gene product and that the activity is essential for transformation of cells to the malignant state. While tight association of traces of a cellular kinase with the native form of pp60^{src} cannot be completely ruled out, they support the conclusion that the *src* gene product is either a protein kinase itself, or functions in close association with a cellular protein kinase.

As well as avian sarcoma virus, protein kinase activity has been found associated with several other transforming gene products, including those from the DNA-containing adenovirus, SV40, and polyoma viruses^{17-21,2}, and from the RNA-containing Abelson murine leukaemia and Fujinami viruses^{3,22}. In each case protein kinase activity is detected in immunoprecipitates of viral proteins using the assay developed by Collett and Erickson. However, phosphorylation in most cases occurs not on the immunoglobulin chain, but on the viral protein itself. Whether this

Thomas Langan is Associate Professor of Pharmacology at the University of Colorado School of Medicine.

tyrosine content of total cell protein during transformation by avian sarcoma virus. When the *src* gene contains a temperature sensitive mutation, phosphotyrosine content rapidly declines to normal values and rises again as cells are shifted from permissive to restrictive temperatures and back²⁵. Not all of the increase in phosphotyrosine can be accounted for by the apparent autophosphorylation of pp60^{src} which occurs in transformed cells, indicating that tyrosine residues in cellular proteins are phosphorylated. Two specific proteins that appear to be substrates of cellular origin are under study. One is a 50,000 MW protein found in immunoprecipitates of pp60^{src} because of its affinity for the kinase protein⁴ and the second is a 34,000 MW protein that is clearly identifiable in normal cells, and which undergoes a transformation dependent and temperature-sensitive phosphorylation in virally infected cells^{26,27}. The 34,000 MW protein, extensively purified from normal chicken cells, is phosphorylated by purified pp60^{src} kinase, and tryptic digests phosphorylated *in vitro* and *in vivo* show the same patterns of tyrosine-containing phosphopeptides in two-dimensional maps²⁸. Determination of the functions of the substrate proteins may not be a simple task and it is possible that pp60^{src} has additional functions which are important for the transformation process.

Tyrosine phosphorylation may not be a universal feature of transformation mechanisms as although transformation

by the related Abelson virus is associated with elevated phosphotyrosine in proteins (Hunter, pers. comm.) in several other viruses including, surprisingly, polyoma and SV40, there is no elevation²⁵. However, the kinases in these cases may be more specific in catalyzing the tyrosine phosphorylation reactions required for transformation, so that no increase in total content is detectable. Two chemically transformed cell lines examined also did not have elevated phosphotyrosine levels.

DNA sequences closely related to the sequences in the avian sarcoma *src* gene are present in normal cells of vertebrates ranging from birds to man²⁹, and transcripts of these sequences are found in polyribosomes³⁰, indicating that the sequences code for a protein with a highly conserved general cellular function. In keeping with this, normal cells also contain a 60,000 MW protein which is structurally related to pp60^{src}, as indicated by immune cross-reaction and by peptide mapping, and which is also associated with protein kinase activity acting on tyrosine residues^{31,32} (Erikson, pers. comm.). The amount of the normal cell protein is some

30-50-fold lower than that of pp60^{src} in transformed cells. These findings suggest that the viral protein may be related to a normal cell protein involved in some universal cellular function such as maintenance of control over cell division. Malignant transformation might result from a qualitative change in the substrates phosphorylated, or from an uncontrolled increase in a normal cell phosphorylation reaction. In either case, investigation of the substrates phosphorylated by the normal cell protein and by any similar kinases seems worthwhile. One interesting candidate has already been identified. When cell membranes from human epidermoid cells bind epidermal growth factor the binding is associated with the phosphorylation of a membrane protein (possibly the receptor for the growth factor) and this phosphorylation occurs predominantly on tyrosine residues³⁴ (S. Cohen, pers. comm.). Future study of this as well as other protein phosphorylation reactions involving tyrosine residues may shed light on important control mechanisms in both normal and malignant cells. □

Repressor represses *cro* represses repressor represses *cro* . . .

from Mark E. Furth

THE discovery of regulatory genes in bacteria and maize quickly invited speculation on how simple control elements such as repressors and operators might determine cell differentiation¹. Are these elements connected in circuits to produce states of gene expression which can be propagated through many cell generations²? Stable regulatory circuits are easy to construct on paper, but more difficult to unravel in the laboratory. The choice between the lysogenic and lytic life-styles of bacteriophage λ has served as one experimental model of a 'developmental switch' between alternative stable states. A regulatory circuit central to this switch has now been explained in exquisite molecular detail³.

The circuit involves two proteins, encoded by genes *cI* and *cro*. Each protein represses the other's synthesis. The *cI* protein (often called the λ repressor) maintains the lysogenic state in which λ replicates passively, its DNA integrated into the chromosome of its bacterial host. In this state the repressor shuts off viral genes, including *cro*, which are required for active growth. The *cro* protein (called *cro*), in contrast, promotes lytic growth by shutting off the repressor gene *cI*. Either of the two stable states of gene expression, *cI* turned on and *cro* turned off, or *cro* turned on and *cI* off, can be maintained for many generations in cells carrying particular λ mutants⁴. These states are regulated by

binding of the repressor and *cro* to an operator called *O_R* which lies between the *cI* and *cro* genes. The bound proteins control the initiation of messenger RNA synthesis at two distinct sites: the promoter *P_{RM}* for transcription of *cI*, and the promoter *P_R* for transcription of *cro* (see figure)^{5,8,12}. How do the repressor and *cro* differentially affect the activity of these two promoters?

One key to the control of *cI* and *cro* expression lies in the fine structure of the operator. M. Ptashne and colleagues discovered several years ago that *O_R* contains three separate binding sites for repressor (*O_{R1}*, *O_{R2}*, *O_{R3}*; see figure)^{7,17}. These sites are closely related, but the nucleotide sequence of each is unique. When *cro* was purified⁹, its target sequences could be compared with those for repressor. Remarkably, although the two proteins show no obvious structural homology^{10,11,18}, they bind to the same three sites in *O_R* (ref. 12). However, the proteins differ in the order in which each occupies these sites. Repressor first binds coordinately to *O_{R1}* and *O_{R2}*, and binds to *O_{R3}* only at much higher protein concentration¹³. *cro* first binds to *O_{R3}*, and then to *O_{R1}* and *O_{R2}*¹². This difference in the relative affinity for binding sites in the operator is critical to gene regulation, because occupation of each site by a repressor or *cro* molecule affects promoter activity in a specific way.

- Collett & Erikson *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021 (1978).
- Eckhart *et al. Cell* **18**, 925 (1979).
- Witte *et al. Nature* **283**, 826 (1980).
- Hunter & Sefton *Proc. natn. Acad. Sci. U.S.A.* **77**, 1311 (1980).
- Collett *et al. Nature* in the press.
- Hanafsa in *Comprehensive Virology* **10**, 401, (ed. Fraenkel-Conrat & Wagner.) (Plenum Press, New York 1977).
- Purchio *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 4661 (1977).
- Purchio *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 1567 (1978).
- Collett *et al. J. Virol.* **29**, 770 (1979).
- Levinson *et al. Cell* **15**, 561 (1978).
- Rubsamén *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 967 (1979).
- Sefton *et al. J. Virol.* **33**, 220 (1980).
- Maness *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 5028 (1979).
- Erikson *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 6260 (1979).
- Erikson *et al. Nature* **274**, 919 (1978).
- Sefton *et al. J. Virol.* **30**, 311 (1979).
- Lassam *et al. Nature* **277**, 241 (1979).
- Tjian & Robbins *Proc. natn. Acad. Sci. U.S.A.* **76**, 610 (1979).
- Griffin *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2610 (1979).
- Smith *et al. Cell* **18**, 915 (1979).
- Schaffhausen & Benjamin *Cell* **18**, 935 (1979).
- Lee *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2018 (1980).
- Giachero & Hager *J. biol. Chem.* **254**, 8113 (1979).
- Adler *et al. J. biol. Chem.* **250**, 6264 (1975).
- Sefton *et al. Cell* (in the press).
- Erikson *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, (in the press).
- Radke & Martin *Proc. natn. Acad. Sci. U.S.A.* **76**, 5212 (1979).
- Erikson *et al. Phil. Trans. R. Soc.* (in the press).
- Spector *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4102 (1978).
- Spector *et al. Cell* **13**, 381 (1978).
- Collett *et al. Cell* **15**, 1363 (1978).
- Oppermann *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 1804 (1979).
- Collett *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 3159, (1979).
- Carpenter *et al. J. biol. Chem.* **254**, 4884 (1979).

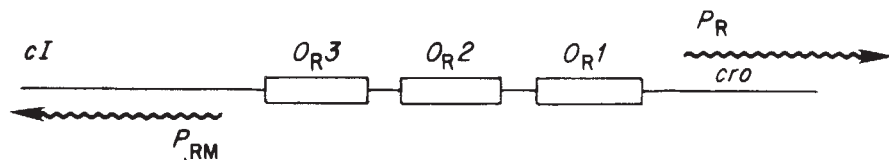


Fig. O_R region of λ DNA. Promoters P_{RM} and P_R direct transcription (wavy arrows) of genes cI (encoding repressor) and cro respectively. Repressor and cro proteins bind to sites O_{R1} and O_{R3} . The distance between mRNA startpoints is 82 base pairs. Each binding site is 17 base pairs long.

B. Meyer, R. Maurer and M. Ptashne have identified the function of protein molecules bound to each site in the operator¹⁴⁻¹⁶. They devised a strategy to measure P_R and P_{PM} activity over a wide range of repressor and cro concentrations in intact cells, and to determine the effects of mutations in the three binding sites. The trick was to place the expression of an easily assayed enzyme, β -galactosidase, under the control of each phage promoter, and to introduce O_R mutations into the fused operons. Levels of the λ regulatory proteins were manipulated by similar genetic sleight-of-hand; the expression of cro or cI was placed under control of an inducible bacterial operon and the rate of synthesis was set by varying the concentration of the inducer. The special elegance of the approach is that it uncouples the synthesis of the regulatory proteins from the promoters which normally control their expression, so that measurements of promoter activity can be made under many steady-state conditions.

The results of studies with the fused operons show that each of the outer operator sites, O_{R3} and O_{R1} , regulates both promoters. Either a repressor molecule or a cro molecule bound to O_{R3} turns off transcription from P_{RM} , but does not turn off transcription from P_R . Conversely, either protein bound to O_{R1} or O_{R2} turns off transcription from P_R but not from P_{RM} ^{14,15}. Repressor, but not cro , bound to O_{R2} has a second effect: it strongly stimulates transcription from P_{RM} ^{5,6,16}.

The different order of binding of cro and repressor to the three sites in O_R , and the particular role of each site in determining promoter activity largely explain the stable states of cI and cro expression. Because cro binds preferentially to O_{R3} , it represses transcription of cI more readily than it represses transcription of its own gene. By contrast, because repressor binds preferentially to O_{R1} and O_{R2} , it turns off transcription of cro more readily than it turns off transcription of cI . Thus, there are concentrations of cro which prevent synthesis of repressor but permit expression of cro , and there are concentrations of repressor which prevent synthesis of cro but permit, and actively maintain expression of cI .

What determines the order of occupation of the binding sites by the

regulatory proteins, and how do bound protein molecules control the initiation of RNA synthesis? Ultimately, it should be possible to answer these questions at the atomic level of resolution. The present studies of molecular mechanisms focus attention on some of the protein-DNA and protein-protein interactions that contribute to gene regulation. A striking and unexpected finding is that cooperative interactions between bound repressor molecules help to govern the pattern of binding to sites in the operator.

Cro binds to the three sites in O_R independently, in the order set by the protein's intrinsic affinity for each site^{12,13}. Repressor presents a more complex picture. Its intrinsic affinity for the isolated

Table 1 Properties of repressor and cro

	Repressor	Cro
Gene	cI	cro
Size of monomer	236	66 amino acids
Active form	Dimer	Dimer
Order of binding to operator sites (wild-type O_R)	$O_{R1} \approx O_{R2} > O_{R3}$	$O_{R3} > O_{R2} \approx O_{R1}$
Cooperativity in binding to DNA	Yes	No

Table 2 Comparison of promoters P_R and P_{RM}

	P_R	P_{RM}
Intrinsic activity	High	Low
Positive control (stimulation)	No*	Yes, by repressor
Primary site for negative control by repressor or cro	O_{R1} , O_{R2}	O_{R3}

(*or weakly by cro : see ref. 8)

Table 3 Action at O_{R2} : Protein bound at O_{R2}

	Repressor	cro
Transcription from P_R :	Off	Off
P_{RM} :	High (stimulated)	Low (unaltered)

O_{R1} sequence is about 15-fold greater than that for O_{R2} or O_{R3} . Nevertheless, on an intact operator it binds O_{R1} and O_{R2} to about the same extent at concentrations at which O_{R3} remains free. Johnson, Meyer and Ptashne attribute the coordinate binding at O_{R1} and O_{R2} to positive cooperative interactions between repressor molecules¹³. They argue that a molecule bound at O_{R1} facilitates the binding of a second molecule at O_{R2} by direct protein contact. When O_{R1} is free, repressor molecules also can cooperate to fill O_{R2} and O_{R3} . However, a molecule bound at the central site does not interact simultaneously with neighbours on both flanks, and at equilibrium the interaction between molecules at O_{R1} and O_{R2} is strongly favoured. A major consequence of

this cooperative binding is to stabilize the state (O_{R1} , O_{R2} bound by repressor, O_{R3} free) in which cro synthesis from P_R is turned off while repressor synthesis from P_{RM} is stimulated. This is the predominant situation in a cell lysogenic for λ .

The capacity of bound repressor to prevent transcription from one promoter while enhancing transcription from another is one of the most striking features of the repressor/ cro regulatory circuit. Both repressor and cro protein appear to shut off transcription simply by blocking access of RNA polymerase to a promoter^{3,6,9,13,15}. However, the mechanism by which repressor boosts transcription from P_{RM} is not obvious. Meyer and Ptashne¹⁶ have shown that repressor must directly effect this stimulation; it is not merely a consequence of turning off transcription from the nearby promoter P_R . They suggest that repressor at O_{R2} interacts with RNA polymerase at P_{RM} to stabilize polymerase binding or to facilitate a subsequent step in the initiation of RNA synthesis^{3,15,16}. The startpoint for transcription from P_{RM} lies only one base pair further from O_{R2} than does the startpoint for transcription from P_R . Yet repressor bound to O_{R2} helps RNA polymerase to initiate at one promoter, but excludes it from the other¹⁵. Whether repressor acts positively or negatively on transcription must therefore critically depend on the precise orientation of bound repressor with respect to the binding site for RNA polymerase. A remaining challenge is to find an exact molecular explanation for the Janus-like behavior of this regulatory protein.

Perhaps a more imposing challenge is to identify genetic regulatory circuits that control development of multicellular organisms, and which can be dissected as neatly as those of λ . One outstanding lesson from the phage model is that the action of each element of such a circuit can be finely tuned and very flexible, so that a single regulator or target sequence may affect gene expression in very different ways. □

- McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **21**, 197 (1956). Monod, J. Jacob, F. *Cold Spring Harb. Symp. quant. Biol.* **26**, 389 (1961).
- Kauffman, S. J. *Theor. Biol.* **44**, 167 (1974).
- Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts T.M. & Sauer R.T. *Cell* **19**, (1980).
- Eisen, H. & Ptashne, M. in *The Bacteriophage Lambda* (ed. Hershey, A.D.) (Cold Spring Harbor Laboratory, New York, 1971).
- Reichardt, L. F. *J. mol. Biol.* **93**, 267 (1975).
- Meyer, B.J., Kleid, D.G. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4785 (1975). Walz, A., Pirrotta, V. & Ineichen, K. *Nature* **262**, 665 (1976).
- Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B.J. & Sauer, R.T. *Science* **194**, 156 (1976).
- Takeda, Y. *J. mol. Biol.* **127**, 177 (1979).
- Takeda, Y., Folkmanis, A. & Echols, H. *J. biol. Chem.* **252**, 6177 (1977).
- Roberts, T.M., Shimatake, H., Brady C. & Rosenberg, M. *Nature* **270**, 274 (1977).
- Hsiang, M.W., Cole, R.D., Takeda Y. & Echols, H. *Nature* **270**, 275 (1977).
- Johnson, A., Meyer B.J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 178 (1978).
- Johnson, A., Meyer B.J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 506 (1979).
- Maurer, R., Meyer, B.J. & Ptashne, M. *J. mol. Biol.* **139**, 147 (1980).
- Meyer, B.J., Maurer, R. & Ptashne, M. *J. mol. Biol.* **139**, (1980).
- Meyer, B.J. & Ptashne, M. *J. mol. Biol.* **139**, (1980).
- Maniatis, T. *et al. Cell* **5**, 109 (1975).
- Saver, R.T. & Anderegg, R. *Biochem.* **17**, 1092 (1978).

Mark E. Furth is in the MRC Laboratory of Molecular Biology, Cambridge.

The problem of defining a substorm

from S.W.H. Cowley

ALTHOUGH the correlation between periods of geomagnetic disturbance and the occurrence of auroras had been known since the time of Celcius (1741), the first scientific reference to magnetospheric substorms seems to be found in the accounts of Herrick (1838) and Olmstead (1856) who noted that auroras often display 'fits of maximum intensity' at intervals of several hours. Interest in these 'auroral fits' and their associated short-lived polar magnetic disturbances waned towards the end of the nineteenth century, and for the next sixty years interest focussed instead on world-wide magnetic storms which develop and decay on time scales of a day or two. It was not until the exploitation of ground networks of all-sky auroral cameras and magnetometers during the International Geophysical Year (IGY, 1957-8) that it was finally realized that the polar 'fits' constituted the building-blocks of world-wide storms. Magnetic storms build up when intense 'fits', now fortunately renamed 'substorms', occur very frequently.

The phenomenology of the auroral substorm described by Akasofu (1964) on the basis of the IGY data is sketched in the accompanying figure, and this was followed by a monograph (1968) in which the evolution of high-latitude phenomena (for example, magnetic disturbance, ionospheric cosmic radio noise absorption, VLF radio emissions) was linked to a common time frame defined by the auroras. Auroral light emission is produced at ~ 100 - 150 km altitudes by 1-10 keV electron bombardment of the upper atmosphere, the electrons originating in the Earth's magnetosphere (the region extending ~ 10 Earth radii on the day side and at least ~ 1000 on the night side containing the Earth's magnetic field and excluding, at least in part, the solar wind plasma). During the early sixties detailed satellite observations in the magnetosphere

out to large distances, but linked to the polar regions by the Earth's magnetic field lines, were only just becoming available, but it was already clear that large changes in the field and charged particle populations occurred, and that the substorm concept provided an organizing framework.

Akasofu's (1964) auroral substorm consisted of two phases, 'expansion' and 'recovery', both lasting 30 min — 1 hour. Onset of expansion is signalled by a sudden brightening (in ~ 1 min), usually near midnight local time, of one of a series of quiet auroral arcs which may have existed for several hours previously (Fig. a). Brightening is followed by rapid poleward motion, forming a bulge which expands also to the east and west (Fig. b). The western end develops a large scale fold or 'surge', and the quiet arcs along which it propagates can brighten just before the surge reaches and engulfs them. In the morning the arcs break up into patches, drifting eastward at speeds ~ 300 m s $^{-1}$. Recovery starts when the bulge reaches its highest latitude, begins to contract and the auroras gradually resume their quiet-time appearance (Fig c). The occurrence of such displays is very variable, but several during one day is not unusual.

The principal magnetic signature is a decrease in horizontal field strength under the auroral bulge due to an ionospheric westward 'electrojet' current of $\sim 10^6$ amps flowing within it, although smaller perturbations are observed down to low latitudes. In addition, a spatially widespread magnetic signal usually appearing as a damped, ~ 1 min period sinusoid is initiated at the onset of expansion. This characteristic signal is known as a Pi-2 and is thought to be excited by rapid changes in plasma flow in the magnetosphere.

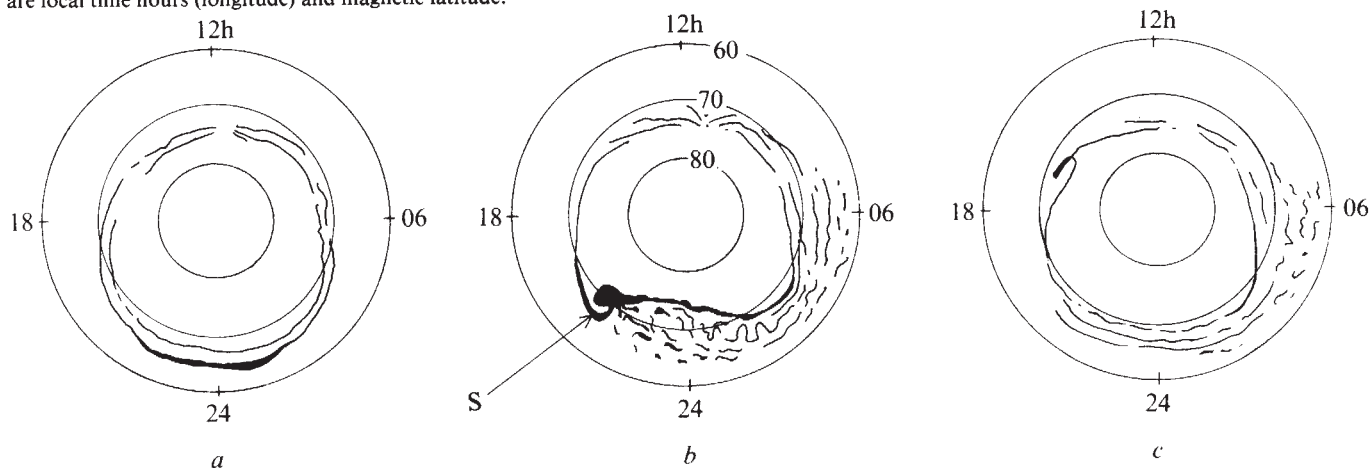
In principle the substorm onset time can be obtained to within a minute or two using either auroral or magnetic data, and such accuracy is often required, for example, to

establish the relative timing of rapidly changing conditions at widely separated locations (for example, ionosphere and night side magnetosphere). In practice, however, timing substorms turns out not to be at all easy, as has become increasingly evident over the past decade, due both to the incomplete coverage provided by ground-based networks and to the increasing realization of their complexity. These problems have encouraged research groups to champion differing signatures as definitions of the substorm time-frame, resulting in a great deal of both scientific and etymological confusion, and providing the basis of much heated argument.

A resolution of these problems was attempted in August 1978, when a three-day workshop was held at Victoria, British Columbia, Canada, attended by nine prominent substorm researchers. Their brief was to attempt to reach consensus on a workable substorm definition, backed by a set of useable signatures. The results of their deliberations have been published very recently (Rostoker *et al. J. geophys. Res.* **85**, 1663; 1980).

The substorm picture discussed above is extended in the new descriptive definition offered by these authors in the recognition of 'multiple onsets', arising from observations of several discrete increases of activity during the expansion phase. Each of these increases is associated with a Pi-2 burst, the formation of a new westward-travelling surge in the auroras, and, as the surge crosses the local meridian, a discrete poleward expansion of the westward electrojet. The first increase in activity is termed the 'onset', and all subsequent increases prior to the start of recovery (defined as before), 'intensifications'. It is noteworthy that their definition includes the necessary appearance of at least one Pi-2 burst, the pulsations being taken to signify that the observed event is of an impulsive nature.

Fig. Sketches of auroras during substorms viewed looking down on the north pole showing (a) thin quiet arcs lying in an oval around the pole with a brightened arc at the equatorward border signalling substorm onset, (b) bulge formation by rapid poleward motion with a westward-travelling surge at its westernmost edge and eastward moving patches in the morning hours, and (c) auroras returning to quiet conditions during recovery. The coordinates are local time hours (longitude) and magnetic latitude.



The second and more important feature of their discussion, however, is concerned directly with the difficulties and ambiguities inherent in currently available ground-based data, and the circumstances required for the drawing of valid conclusions regarding the start of the expansion and recovery phases. It is pointed out, for example, that a single all-sky camera sees only a small part of the whole auroral display at any one time, and as a result the brightening of arcs just to the west of a surge followed by passage of the surge across the local meridian can easily be mistaken for a local onset. In the future any substorm onset time proposed on the basis of local auroral brightening should therefore, the authors say, be substantiated by showing that at the least no brightening

S.W.H. Cowley is an SRC Advanced Fellow in the Blackett Laboratory, Imperial College, London.

occurred a few minutes earlier to the east of the station. Similar treatment is given to the magnetic signature, and it is finally uncompromisingly recommended that researchers use original magnetograms from "as many relevant observatories as possible" to determine individual substorm development. Standard magnetic indices, in common use at present, which are synthesized from a few auroral zone stations are to be eschewed.

Even if the above laborious practices are followed, however, not all onset and recovery times can be determined with confidence due to gaps in ground coverage. Indeed some substorms, particularly those of small amplitude which tend to occur at higher than normal latitudes, may not be detected at all. These inadequacies were particularly exposed in the last decade when the first whole auroral oval pictures became available via the imagers flown by

the US Air Force Geophysical Laboratory on DMSP satellites and by Canadian experimenters on ISIS-2. In these systems the picture is built up on a line-by-line basis as the satellite traverses the polar regions at $\sim 1,000$ km altitudes, so that only one 'snapshot' (of the northern auroral zone) is obtained per orbit (~ 90 min). These images allowed a great step forward in understanding auroral morphology, but were, of course, unable to follow the changes during substorms. A radical solution to the problems emphasized by Rostoker *et al.* is therefore clear, although technically difficult and rather expensive. What is needed, as the authors point out, is a satellite auroral TV system with a resolution better than 10 km spatially and 1 minute in time, which can return images continuously for a period of several hours. Only with such data will the full story of the dynamic auroras be finally revealed. □

Radar measurements of estuarine currents and tides

from E.D.R. Shearman

AN important step in the development of a new remote sensing system is reached when the research moves from comparisons with ground truth to the provision of area surveillance data in a form directly comparable with theoretical models. This point has now been reached for surface current measurement by High Frequency (HF) radar in the dekametric band as described by A.S. Frisch and B.L. Weber¹.

The discovery from which HF radar oceanography has grown was made in 1969 by D.D. Crombie². In observations with a vertically-polarized 13.5 MHz radar, directed seawards from a coastal site, he detected Doppler-shifted echoes with a strong spectral line component at a shift of 0.376 Hz. This shift corresponds to a radial velocity which is that of gravity waves on the sea surface of wavelength 11 m, just half the radio wavelength. There is a selective scattering process, now known as Bragg scattering by analogy with the diffraction of X rays from the regular rows of atoms in a crystal lattice. This process sifts out from the chaotic superposition of waves of different lengths and different directions on the sea-surface those waves travelling radially towards or away from the radar and of wavelength half the radar wavelength.

By using a number of radar frequencies and by directing the radar beam in different directions it thus becomes possible to measure the wavelength spectrum and the directional spectrum of sea waves, whether wind-driven waves or swell. This has led to

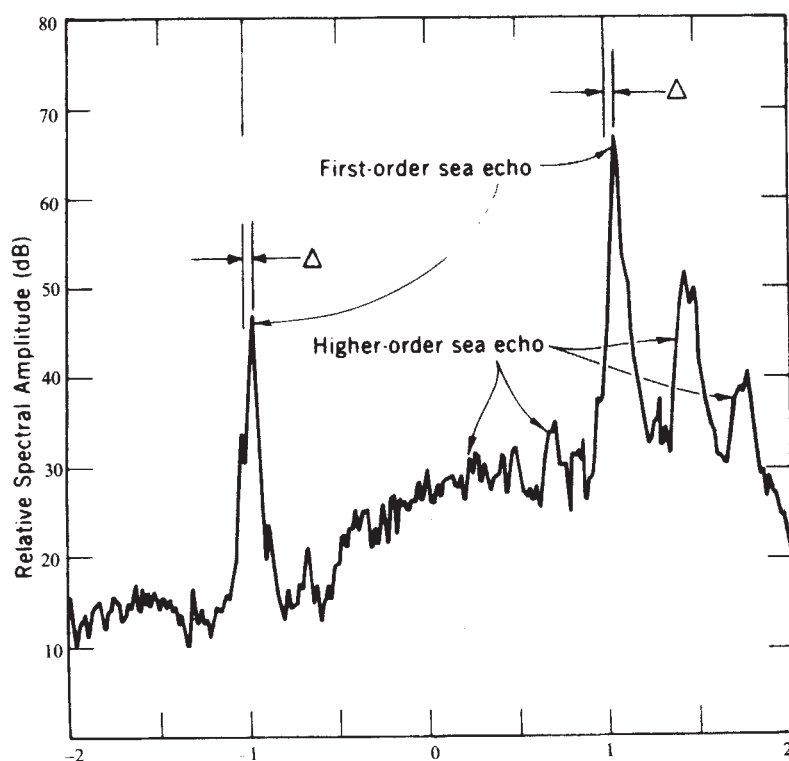


Fig. 1 Example of 9-sample average of 200 second HF ground-wave sea-echo power spectra at 9.4 MHz. Doppler frequency (horizontal axis) is normalised with respect to the expected position of the first order Bragg resonant peaks (here 0.313 Hz). Δ is the normalized shift of the record due to a surface current, as utilized in the NOAA current-measuring radar. Note the narrow Bragg peaks as seen with a narrow-beam radar. (After Barrick *et al.*)

E.D.R. Shearman is in the Department of Electronic and Electrical Engineering, University of Birmingham.

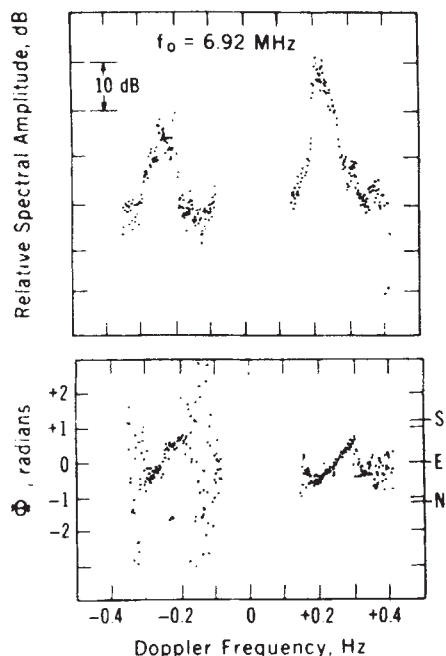


Fig. 2 Upper: Doppler spectrum of current-smear sea echo from the Gulf Stream received on a wide-beam antenna.

Lower: Phase difference between two spaced antennas versus Doppler-shift, demonstrating that echoes from different directions are sorted by current-produced Doppler and can then be identified. (After Crombie *et al.*¹¹).

the remote measurement not only of wave spectra, but also, by inference, of sea-surface wind direction³ and wind speed⁴.

Theoretical investigations by Hasselmann⁵ and by Barrick⁶ make it possible to deduce the wavelength and the directional spectra of sea waves by inversion of second order features in the radar echo spectra from a single patch of ocean at a single radio frequency. The spectrum of Fig. 1, obtained at San Clemente Island, California⁷, illustrates these first and second order features. These techniques already promise a capability for characterising sea-state over a large area of ocean. Crombie's coastal-sited 'ground-wave radar' technique can give coverage up to some 300 km with suitable power and frequency; 'sky-wave radar' using ionospheric refraction covers the area between 1,000 km and 3,000 km from the radar^{8,9}. A useful review of the basic concepts of HF radar oceanography has been given by Barrick¹⁰.

The paper by Frisch and Weber makes use of a different phenomenon from those employed in the above wave and wind measurements. In Fig. 1 it will be seen that the approach and recede first-order Bragg resonant frequencies, nominally at ± 1 on the normalized Doppler frequency scale, are displaced from symmetry about zero, as if the approaching waves were travelling faster towards the radar than the receding waves away from the radar. This is what would be expected if the water surface were being carried towards the radar by the radial component of a tidal stream or

current. (In the case of sky-wave propagation, such shifts can also be introduced by ionospheric motion, but this is not so for ground-wave radar).

If we consider a ground-wave radar located on the sea coast with an omnidirectional antenna, so that sea areas in all directions contribute to the received echo, a 'current-smearing' will occur. Returns from different directions will be Doppler-shifted by the current to different extents. Fig. 2 shows the results of an experiment done under conditions¹¹ in which reception took place on two spaced antennas. At the top is seen the current-smear spectrum for the approach and recede Bragg resonant waves, while at the bottom is the phase difference between the two antenna signals plotted as a function of Doppler shift. This phase difference, Φ , would be expected to be given by $k_0 d \sin \alpha$ where d is the (known) spacing between the two receiving antennas, α the azimuthal angle relative to the normal to d , and k_0 is the (known) radar wave number. We see that by Doppler analysing the echoes, returns from each direction have been sorted by frequency and that it is then

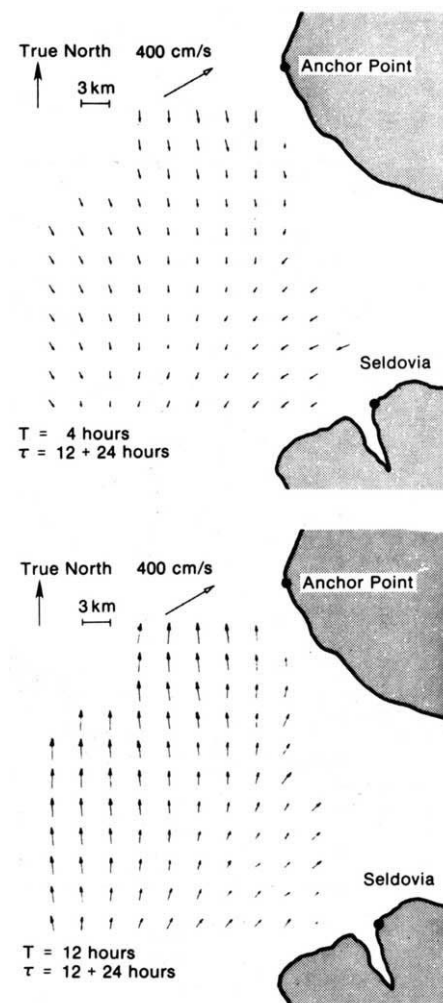


Fig. 3. Two maps depicting the tidal currents in the vicinity of Kachemak Bay at times 8 hrs apart. The current vectors are plotted at 3 km intervals relative to the southern radar site near Seldovia.

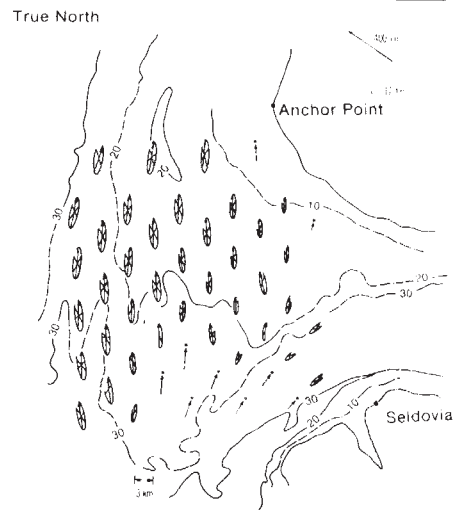


Fig. 4. Tidal ellipses for the 12 hour tidal velocity component in the vicinity of Kachemak Bay, starting at 00 hours on July 13, 1977. The arrow points to the current direction at 00 hours.

possible to measure their direction of arrival by simple relative phase measurement, at separated antennas. This is the principle of the NOAA current measuring radar¹².

Frisch and Weber describe the deployment of two such radars in the Cook Inlet, Alaska. Fig. 3 shows a snapshot of current vectors obtained by putting together the radial current components measured from the two sites, Seldovia and Anchor Point. The data was analysed by extracting from the radar results at grid locations spaced 3 km apart, the North-South and East-West components of the current. The values of these components were available from observations taken every fifteen minutes throughout twenty-four hours. A least square fit was then carried out to estimate the nominal twelve and twenty-four hour tidal components for the surface currents as a function of position. The current velocity at the grid locations showed a considerable but orderly amplitude and phase variation over the radar coverage area.

A compact summary of the kind of results that can be obtained by this method is given in Fig. 4, here showing the tidal ellipses for the 12 hr component of tidal velocity. □

1. Frisch, A.S. & Weber, B.L. *J. geophys. Res.* **85**, 485 (1980).
2. Crombie, D.D. *Nature* **223**, 1325 (1969).
3. Long, A.E. & Trizna, D.B. *IEEE, Trans. Ant. Prop.*, **AP-21**, 680 (1975).
4. Barnum, J.R., Maresca, J.W. & Serebeny, S.M. *IEEE, Trans. Ant. Prop.* **AP-25**, 128 (1977).
5. Hasselmann, K. *Nature Phys. Sci.* **229**, 16 (1971).
6. Barrick, D.E. *Remote Sensing of Environment*, **6**, No. 2, 201 (1977).
7. Barrick, D.E., Headrick, J.M., Bogle, R.W. & Crombie, D.D., *Proc. IEEE*, **62**, No. 6, 673 (1974).
8. *Proc. IEEE. Special issue on modern radar technology and applications*. *Proc. IEEE* **62**, No. 6 (1974).
9. Shearman, E.D.R., Sandham, W.A., Bradley, E.N. & Bradley, P., *AGARD Conf. Proc.* **263**, 30 (1979).
10. Barrick, D.E. *Remote Sensing of the Troposphere* (ed. Derr, V.E.) Chap. 12 (US Govt. Printing Office, Washington 1972).
11. Crombie, D.D., *et al.* 1970. Reproduced in NOAA Technical Report ERL 373, WPL 47, (U.S. Dept. of Commerce 1976).
12. Barrick, D.E., Evans, M.W. & Weber, B.L. *Science* **198**, 138 (1977).

ARTICLES

Nd–Sr isotopic relationship in granitoid rocks and continental crust development: a chemical approach to orogenesis

Claude J. Allègre & Dalila Ben Othman

Laboratoire de Géochimie et Cosmochimie, Institut de Physique du Globe, Département des Sciences de la Terre, Universités de Paris VI et VII, 2 Place Jussieu, 75230 Paris Cedex 05, France

Nd and Sr isotopic compositions of young granitoid rocks from diverse locations show that none of them is derived from an oceanic mantle. Instead a large amount of recycled older continental crust must be involved in their genesis. In contrast old granitoids (older than 2,000 Myr) seem to have initial isotopic ratios of Nd and Sr quite close to the planetary mantle evolution line. These observations are interpreted as evidence of a quasi-continuous continental growth through geological time. We propose a mixing model to explain quantitatively the generation of granitoids from mantle and recycled continental materials.

THE process of orogeny incorporates a variety of stages such as folding, metamorphism and granitization, and its main chemical result is the formation of a piece of continental crust. As granitoids are the main components of continental crust, the problem of their genesis is linked to that of the origin and evolution of continents. Two extreme theories have been proposed:

- (1) continental crust is created from the mantle by a continuous differentiation process throughout geological time;
- (2) continental crust has been differentiated once from the mantle, early in the Earth's history, since then this material has been continuously recycled into the new crust.

The formation of granites can also be described by two models: either as the ultimate product of magmatic differentiation, or as a product of crustal anatexis^{1,2}. Although using Sr or Pb isotopes as petrogenetic tracers has clarified some problems²⁻⁷ no widely accepted model has been proposed so far.

Following the work by Richard *et al.*⁸ Nd isotopes were used for extremely powerful tracer studies⁹⁻¹³. Their usefulness is due to two geochemical properties of rare-earth elements (REEs): they are not very mobile and large geochemical fractionation of the parent/daughter Sm/Nd ratio is difficult.

This article considers the application of the coupled Nd–Sr isotopic systems to the genesis of continental crust: old and young granitoids are compared.

Petrogenesis of recent granitoids

We have investigated the Nd systematics in granitoids younger than ~350 Myr (the age correction for ¹⁴⁷Sm *in situ* decay is, therefore, rather small) representative of the various geodynamic environments: (1) island-arcs (Japan); (2) continental subduction zones (Andes); (3) complex subduction areas (southern California Batholith, Sierra Nevada); (4) continental collision zones (Himalayan and European Hercynian granites). Results for the major petrographic types of granitoids are given in Table 1 and illustrated in Figs 1 and 2.

Subduction zones

Japan. We have studied two typical Japanese granitoids from Okue and Owe in the centre of Honshu Islands, which are 21- and 12-Myr old respectively. Like the whole calc-alkaline suite, these granitoids are associated with the general subduction process. $\epsilon_{Nd} = -3.3$ and -3.1 (see Table 1 for a definition of ϵ_{Nd}). Because these values are significantly lower than those for mid-oceanic⁸⁻¹¹ and island-arc basalts^{12,13}, direct derivation from a downgoing plate is ruled out. These granitoids are at least partly composed of crustal material, which may have been

incorporated into the subducted plate as detrital sediments, or it might result from remelting of the overlying continental crust induced by the presence of hot basic or calc-alkaline magmas. **California.** Two typical batholiths occur in California: in southern California and in the Sierra Nevada. Although the tectonic situation at the time of their formation is not clear, it is considered to be related to subduction processes¹⁴.

The southern California batholith is a complex which includes many different types of granitoids with ages ranging from 80 to 103 Myr on the east and 105 to 130 Myr on the west¹⁴. Detailed studies by Silver *et al.*¹⁴⁻¹⁷ using Sr and Pb isotopes and trace elements, and by Taylor and Silver¹⁸ using ¹⁸O/¹⁶O and ⁸⁷Sr/⁸⁶Sr ratios, have shown that the batholith can be divided into a western and an eastern part, the former being older and emplaced at a shallower depth than the latter.

The variation of ¹⁸O/¹⁶O and ⁸⁷Sr/⁸⁶Sr seems to confirm that several sources are involved in the genesis of this batholith. The value $\epsilon_{Nd}^i = +3.6$ for the San Marcos gabbro is consistent with that of mid-oceanic ridge basalts (MORB) contaminated by some crustal material. The Rubidoux granite, a 'charnockitic granite' has $\epsilon_{Nd}^o = -1.4$, which is close to the present chondritic value, while for the Lakeview Tonalite $\epsilon_{Nd}^o = -9.9$ which is very different from any mantle material value and implies a large crustal component.

In the study of the Sierra Nevada batholith by Hurley *et al.*¹⁹ using the ⁸⁷Rb/⁸⁷Sr method, some isotopic heterogeneities were pointed out within the batholith. More recent studies²⁰ gave a better definition of this heterogeneity: this batholith is composed of many plutons almost continuously emplaced from 134 to 85 Myr and with some older ones around 160 Myr. In the southern California batholith there is a systematic variation from west to east for both Sr and Pb isotopes. Note that the four rocks studied here clearly do not constitute a comprehensive survey of the two massives. As the ϵ_{Nd}^o vary between -7 and -3.3 and is clearly outside any mantle trend, this implies the participation of crustal material in the genesis of these batholiths. Again, none of these granitoids can be considered as the direct product of mantle type material—remelting of old continental crust is needed. This agrees with the studies of Magaritz and Taylor and Armstrong *et al.*^{21,22}. Both batholiths are heterogeneous and seem to contain oceanic material mixed with previous continental crust.

Peru. The tectonic environment in the Andes is important because it has been considered as the archetypal continental growth mechanism. Although we have studied only one granitoid, more extensive research has been carried out on the different types of lavas which are widely held to be the extrusive

equivalent of true granitoids. We have studied two andesites, three rhyolites and one dacite, from the Barroso and Arequipa volcanic fields in southern Peru. Sr and Pb isotope studies have shown that these rocks have been contaminated by continental crust^{23,24}. This contamination seems to be differential and more extensive than for other samples. The values of ϵ_{Nd}^i are all negative between -1.4 and -5.1; none of these values are thus consistent with a pure derivation from oceanic mantle. These results agree with those of De Paolo and Wasserburg⁹ which are based on the analysis of one andesite and imply a crustal contamination.

The continental collision situation

The continental collision environment must exist in Himalayan granites and has also been proposed to account for the French hercynian segment²⁵.

The Himalayan granitoids can be divided into two categories: those which are situated along the Main Central thrust (MCT), a gigantic intracontinental thrust inside the Indian shield^{25,26} and those which are on the edge of the Tsanpo suture, composing the Laddakh batholith.

The MCT granites are all leucogranites (aluminous, two mica alkaline granites). An age determination of 29 Myr for the Manaslu massif has been questioned²⁷, and the other Palung Daman and Makalu-Kakani granitoids seem to be slightly different and not well dated either²⁸. However, for general stratigraphic reasons, the ages of these granites are taken as between 50 and 15 Myr (ref. 26) which gives initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of around 0.780. Such high ratios clearly represent the remelting of old continental material, which agrees well with the petrological study by Lefort²⁶. As expected, ϵ_{Nd}^i values are strongly negative, from -10.3 to -15.8 and are the lowest values ever reported for young granitoids. There is little doubt that these values, like those for Sr, indicate the almost pure remelting of old continental crust.

The Laddakh complex is a large granodiorite complex which represents the granitoids associated with the major collision between India and Eurasia. Two samples have been analysed with ϵ_{Nd} values of +1.8 and -1.4. Both values may be interpreted as a record of material derived from an oceanic crust contaminated by crustal material (as for California or Japan).

The 'pyrenean' granitoids have been extensively studied from the chronological, isotopic, and petrological points of view²⁹⁻³⁴. They were deposited during the hercynian orogeny and can be divided into two groups. The first is represented by early granites emplaced 320 Myr ago with a high initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio (0.712) characteristic of many gneissic xenoliths; a typical example is the Aix-les-Thermes 2-mica granite. The second group is younger (280 Myr) and has a lower initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio²⁹. These granitoids are characterized by xenoliths of quartz dioritic composition and a complex set of different facies³². The Querigut granite³³ and the Costabonne granite belong to this second group.

According to Mattauer's interpretation of the hercynian orogeny in western Europe³⁴, these granites were generated during a continental collision very similar to the Himalayan one. All of the granites have negative ϵ_{Nd}^i values between -0.7 and -8.9 which again cannot be explained by direct derivation from the oceanic mantle. This conclusion is supported by trace element results³⁵.

Nd-Sr relationship on modern granitoids

The relationship between the Sr and Nd isotopic ratios of modern basalts suggested by Richard *et al.*⁸ and demonstrated by others^{10,11} is a very important and distinctive signature of mantle evolution³⁶. It seems that the initial ratios of Sr and Nd of any material directly derived from the mantle would probably plot on this general trend. On the other hand, as we have already emphasized, the studies of Carter *et al.*³⁷ have shown that the Sr-Nd isotope diagram can clearly discriminate a mantle type origin from a lower or upper crustal origin. Therefore, the Nd-Sr diagram should give a clear distinction between different hypotheses of granitoid genesis. Let us consider the general trends in ϵ^o notations. The present terrestrial mean corresponds to $\epsilon_{\text{Nd}}^o = 0$ ($^{143}\text{Nd}/^{144}\text{Nd} = 0.51264$) (see above) and by definition $\epsilon_{\text{Sr}}^o = 0$ ($^{87}\text{Sr}/^{86}\text{Sr} = 0.70478$) (ref. 36). The oceanic mantle, and probably some of the subcontinental mantle plot in the upper left quadrant ($\epsilon_{\text{Nd}}^o > 0$ and $\epsilon_{\text{Sr}}^o < 0$). In this quadrant there is a negative correlation between ϵ_{Nd}^o and ϵ_{Sr}^o . Carter *et al.*³⁷ have shown that Lewisian granulites normally plot in the lower left quadrant ($\epsilon_{\text{Nd}}^o < 0$ and $\epsilon_{\text{Sr}}^o < 0$) although a few granulites do have positive ϵ_{Sr}^o . The old granitoids or gneisses in amphibolite facies

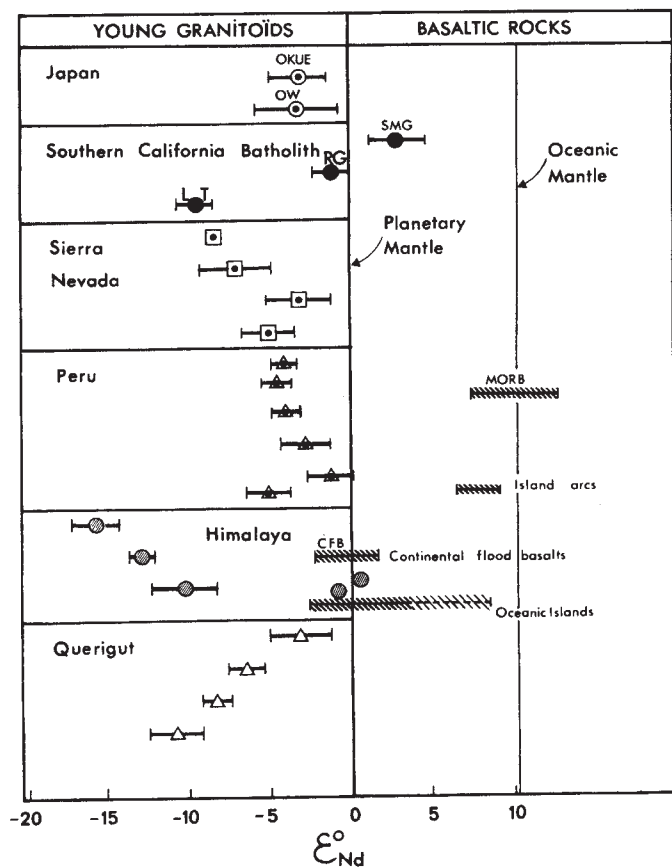


Fig. 1 Raw values of young granitoids expressed in ϵ_{Nd}^o . The reference values for MORB, oceanic islands, continental flood basalts and island arcs are from refs 8, 9-11, 37: ○, Japan; ●, Southern California Batholith; □, Sierra Nevada; ▲, Peru; ●, Himalaya; △, Pyrenees, Querigut; ●, Costabonne; ◆, Aix les Thermes; ▲, In'ize. We adapted the technique of Richard *et al.*⁸ to the granite problem. To dissolve all the accessory minerals (such as sphenes and zircons) often rich in REE, they were attacked in a Teflon bomb under pressure (see ref. 73), HNO_3 and HClO_4 were added, reevaporation took place and they were then dissolved in 6 M HCl. Measurements were made using the triple filament technique. The total blank is about 40 pg and is, therefore, negligible in our study. Notations are defined by:

$$\epsilon_{\text{Nd}}^o = \frac{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{sample, Today}} - (^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, Today}}}{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, Today}}} \times 10^4$$

$$(^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, Today}} = 0.51264 \text{ (ref. 74).}$$

$(^{143}\text{Nd}/^{144}\text{Nd})_{\text{sample, Today}}$: measured value of the sample. $\epsilon_{\text{Nd}}^{\text{initial}}$ for a sample of age T_1 is defined in the same way:

$$\epsilon_{\text{Nd}}^i = \frac{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{sample, } T_1} - (^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, } T_1}}{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, } T_1}} \times 10^4$$

$$(^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, } T_1} = (^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, Today}} - (^{147}\text{Sm}/^{144}\text{Nd})_{\text{Pla}} (e^{\lambda T_1} - 1);$$

$$(^{143}\text{Nd}/^{144}\text{Nd})_{\text{Pla, } T_1} = (^{143}\text{Nd}/^{144}\text{Nd})_{\text{sample, Today}} - (^{147}\text{Sm}/^{144}\text{Nd})_{\text{sample}} \cdot (e^{\lambda T_1} - 1).$$

The Sr isotopic composition has been measured in most samples following the technique described elsewhere⁷⁵. For Sr, ϵ_{Sr}^i can also be defined in the same way³⁶.

Table 1 Major petrographic types of the granitoids

Samples	$^{147}\text{Sm}/^{144}\text{Nd}$	$^{143}\text{Nd}/^{144}\text{Nd}^*$	ϵ_{Nd}^0	ϵ_{Nd}^i	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{87}\text{Sr}/^{86}\text{Sr}_{\text{initial}}$	$^{87}\text{Rb}/^{86}\text{Sr}_{\text{initial}}$	Model Sr age (Myr)	Model Nd age (Myr)	Real Age (Myr)	ϵ_{Nd}^i	ϵ_{Sr}^i
Young granitoids												
Japan Okue 2		0.51247 ± 9	-3.3	-3.3	0.7062 ± 1	0.70610	2.15	49		21	-3.3	+11.3
Owe 2		0.51248 ± 13	-3.1	-3.1	0.7077 ± 1	0.70713	3.00	72		12	-3.1	+34.4
Himalaya												
Nepal 22 D. Manaslu	0.141	0.51183 ± 7	-15.8	-16.3	0.74332 ± 40§	0.7408§	5.98§	471	2,340	28	-16.3	+513.6
Palung A 404		0.51197 ± 4	-13.1	-13.1	0.82131 ± 50§	0.8000§	37.8§	223			-13.1	+1,456.0
Makalu 132	0.121	0.51211 ± 10	-10.3	-10.10					1,110	26	-10.10	
Laddakh												
granodiorite HB68	0.108	0.51256 ± 3	-1.4	-1.4	0.7050 ± 1§	0.7045	0.170	190	65	51	+2.0	2.66
diorite HB74	0.617	0.51273 ± 3	+1.8	+1.8	0.7048 ± 1§	0.7045	0.352				-1.1	2.28
Southern Batholith of California												
San Marcos G. 60-5	0.153††	0.51279 ± 5	+2.9	+3.6	0.7040	0.7032	0.38	-190	-565	120	+3.6	-18.2
Rubidoux G. EL 38		0.51257 ± 5	-1.4	-1.4	0.7121	0.7045	4.42	121		120	-1.4	-1.0
Lakeview Tonalite 60-2		0.51213 ± 4	-9.9	-9.9	0.7058	0.7045	0.73	114		120	-9.9	-1.1
Sierra Nevada												
R 5075		0.51228 ± 11	-7.0	-7.0	0.7072 ± 2	0.7063	0.68	293		90	-7.0	+23.6
R 5084		0.51247 ± 10	-3.3	-3.3	0.7106 ± 1	0.70937	1.13	404		90	-3.3	+64.4
R 5079		0.51237 ± 8	-5.3	-5.3	0.7106 ± 1	0.70950	0.97	474		90	-5.3	+66.6
R 4602	0.107	0.51220 ± 2	-8.6	-6.6	0.7083 ± 2	0.7063	0.79	361	775	180	-6.6	+24.4
Peru												
Granite P 293		0.51238 ± 10	-5.1	-5.1	0.7235 ± 1	0.7105	4.02	342		230		83.4
Rhyolites												
P 255	0.094§§	0.51248 ± 8	-3.1	-3.1	0.70705 ± 5#	0.7070 ± 3#	1.95#	87	246	1.74	-3.1	32.2
P 293	0.116§§	0.51257 ± 4	-1.4	-1.4	0.73928 ± 7#	0.7174 ± 3#	909.0#	2.8	138	1.74	-1.4	171
P 292	0.138§§	0.51238 ± 6	-5.1	-5.1	0.7484 ± 6#	0.7174 ± 3#	1172.4#	2.6	715	1.74	-5.1	206.1
Dacite												
P 254	0.086§§	0.51243 ± 3	-4.1	-4.1	0.70619 ± 6#	0.70618 ± 6#	0.57#	209	298	1.74	-4.1	20
Andésites												
P 290	0.104§§	0.51242 ± 4	-4.3	-4.3	0.70551 ± 5#	0.70551 ± 6#	0.54#	108	375	1.74	-4.3	10.3
P 289	0.085§§	0.51239 ± 1	-4.9	-4.9	0.70574 ± 4#	0.70574 ± 6#	0.54#	151	352	1.74	-4.9	13.6
French Hercynian granites												
From Querigut												
Quartz diorite QT 08	0.137‡‡	0.51250 ± 9	-2.7	-0.7	0.7085 ± 9	0.70633	0.55	578	378	280	-0.7	26.7
Monozonitic granite QT 12	0.097‡‡	0.51230 ± 6	-6.6	-3.2	0.7214 ± 2	0.7101	2.89	426	540	280	-3.2	77.7
Granodiorite QT 52	0.112‡‡	0.51221 ± 3	-8.4	-5.5	0.7136 ± 7	0.7099	1.09	632	806	280	-5.5	68.6
Calcaline granite with												
bitotites QT 15												
Granite Costabonne	0.142	0.51239 ± 45	-4.9	-3.1	0.7230¶¶	0.711	3.03	445	741	265	-3.1	104.4
Aix les Thermes AX4	0.110	0.51204 ± 60	-11.7	-8.2	0.7829¶¶	0.711	15.9	355	1,100	325	-8.2	92.6
Sahara												
Inize	0.1197	0.51170 ± 5	-18.3	-12.7					1,930	600	-12.7	
Precambrian granitoids												
Grenville												
5208 Westport Granite	0.0920	0.51209 ± 3	-10.7	+2.5	0.753**	0.704**	3.25	1,090	835	1,020	+2.5	+43.4
5210 Blue Mountain	0.1325	0.51210 ± 9	-10.4	-0.37	0.727	0.7041 ± 4	1.30	1,290	1,330	1,285	-0.37	+5.1
Rapakivi Parguasa (Venezuela)	0.092	0.51160 ± 2	-20.3	-0.35	0.8405	0.7004 ± 1	6.53	1,500	1,510	1,530	-0.35	-39.4
Tassendjanet												
L 418	0.085	0.51125 ± 3	-27.1	+2.0	0.713 ± 1	0.7018	0.39	1,950	1,930	2,090	+2.0	-10.3
M 697	0.086	0.51123 ± 1	-27.5	+1.0					1,980	2,090	+1.0	
Ivory Coast Granite Cl ₂	0.107	0.51150 ± 6	-22.3	+0.1	0.7141 ± 6	0.7010 ± 5	0.416	2,010	2,000	2,010	+0.1	-0.0
Swaziland												
S ₁ , S _A	0.115	0.51098 ± 2	-32.3	+2.1					3,200	3,400	+2.1	
S ₁ , S _B	0.106	0.51066 ± 2	-38.6	-0.2					3,400	3,400	-0.2	
China granites												
CHID	0.119	0.51150 ± 3	-22.3	+0.8	0.7033 ± 8##	0.7025 ± 5##	0.023##	1,590	2,320	2,400	0.8	+18.9
CHIB	0.068	0.51073 ± 5	-37.3	+0.9	0.7165 ± 7##	0.7025 ± 5##	0.417##	2,550	2,340	2,400	0.9	+8.6

* Normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$; errors are 2σ mean.

† ϵ_{Nd}^0 and ϵ_{Nd}^i relative to Juvinas; $^{143}\text{Nd}/^{144}\text{Nd}$ today = 0.51264; $^{147}\text{Sm}/^{144}\text{Nd}$ today = 0.1936; $\lambda^{147}\text{Sm} = 6.54 \times 10^{-12} \text{ yr}^{-1}$.

‡ $^{87}\text{Rb}/^{86}\text{Sr} = 1.39 \times 10^{-11} \text{ yr}^{-1}$; $^{87}\text{Sr}/^{86}\text{Sr}$ today = 0.70478; $^{87}\text{Rb}/^{86}\text{Sr}$ today = 0.087.

§ Ref. 27. || Unpublished data. ¶ Ref. 28. ** Ref. 23. *** Ref. 55. †† Ref. 12. ‡‡ Ref. 34. §§ Dupuy, unpublished data. |||| Ref. 8. ¶¶ Ref. 76. ## Feldspar from CH1D and CH1B.

A number of ratios of $^{87}\text{Sr}/^{86}\text{Sr}$, taken from the literature (for example, Aix les Thermes, and Costabonne) are given here with smaller number of digits (3 or 4) and without error bar, due to their ancient determination.

($^{87}\text{Sr}/^{86}\text{Sr}$) initial ratios for Japan, Sierra Nevada, Querigut and part of Peru samples have been calculated from ages determined with other radiochronometers and with the $\lambda_{\text{Rb}} = 1.39 \times 10^{-11} \text{ yr}^{-1}$. Ages for all the other samples have been determined from corresponding isochrons.

and old schists, which might be attributed to the upper or medium continental crust, plot in the lower right quadrant ($\epsilon_{\text{Nd}}^0 < 0$, $\epsilon_{\text{Sr}}^0 > 0$).

To compare the position of modern granites with that of other rock types more quantitatively, we have to remember that ϵ_{Nd}^0 and ϵ_{Sr}^0 values for the old granitoids and gneisses which form the basement in most continental areas strongly depend on their ages. Statistically the older they are, the lower their ϵ_{Nd}^0 and the higher their ϵ_{Sr}^0 . This precludes the possibility of getting a simple field in a ϵ_{Nd}^0 - ϵ_{Sr}^0 diagram for all types of continental crust and therefore, we must consider each case individually.

In the (ϵ_{Nd}^i , ϵ_{Sr}^i) correlation diagram we shall consider the mantle correlation band and the quadrant defined by Carter *et al.*³⁷. Plotting our results for young granitoids, we see that (Fig.

3) most modern granitoids plot in the lower right quadrant, roughly indicating that they show isotopic affinities with the upper or medium crust rather than with the granulite facies. Only the two granitoids of the southern California batholith plot in the lower left quadrant. These massifs may be derived from lower granulite crust (see ref. 37). Thus the correlation diagram qualitatively suggests that no young granitoid is an uncontaminated product of the mantle, and that they all contain some crustal recycled component. We can now study the mixing of a mantle component and a recycled crustal component more quantitatively. Such a theory was first proposed in 1965 (ref. 38) using isotopic evidence, but it can now be modelled.

The equations show (see Fig. 4b) that the mixing depends on three groups of parameters: (1) the isotopic compositions of the

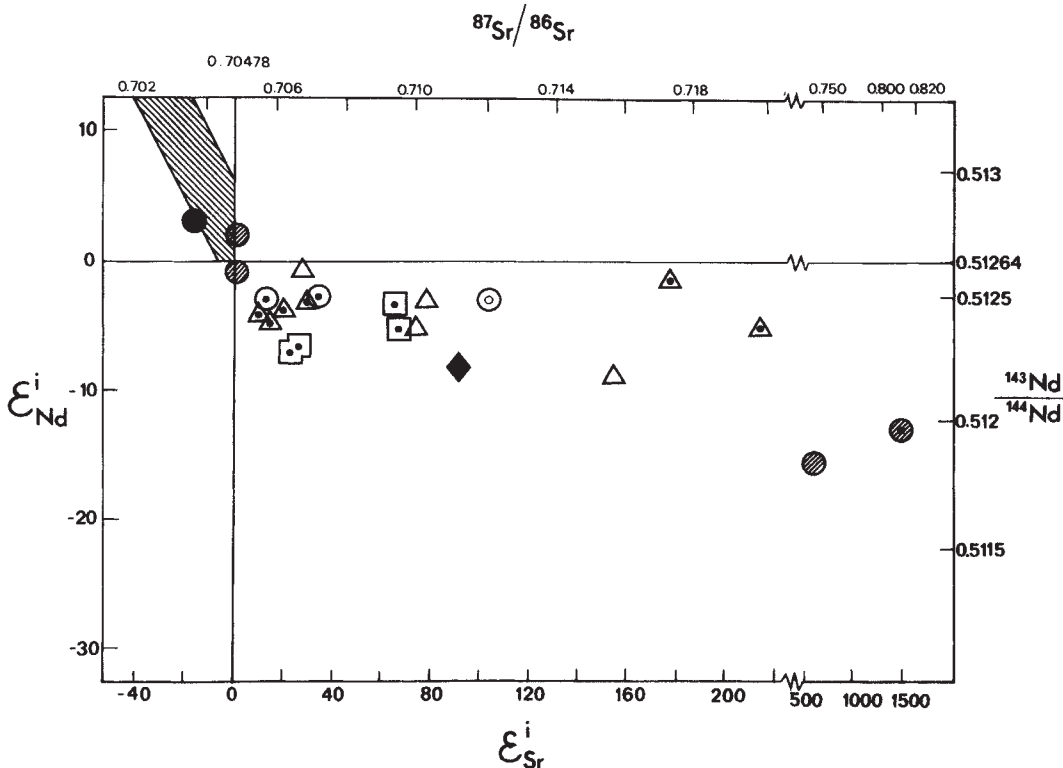


Fig. 2 Comparison of ϵ_{Nd}^i with ϵ_{Sr}^i for young granitoids. Symbols are as in Fig. 1.

two end-members, the mantle and the crustal component; (2) the ratios of Nd and Sr concentrations in crust and mantle, $C_{\text{Nd}}^c/C_{\text{Nd}}^m$ and $C_{\text{Sr}}^c/C_{\text{Sr}}^m$; (3) the mass ratio of the two components, m^c/m^m .

We will select two different points for the isotopic composition of each component. We assume that the mantle component has the composition of a typical oceanic MORB mantle or that it has a planetary composition. For the crustal component we will fix the Sr isotopic ratio as $^{87}\text{Sr}/^{86}\text{Sr} = 0.720$, and take a 'young' crustal component with $^{143}\text{Nd}/^{144}\text{Nd} = 0.5125$ and an 'old' crustal component with $^{143}\text{Nd}/^{144}\text{Nd} = 0.5108$. These limits allow the observed results to be discussed in a general framework.

The mixing equations also involve the Nd and Sr crust/mantle concentration ratios. The values for both the upper and lower crustal component concentrations fall in a fairly narrow range^{37,39,40}. The mantle component concentrations are more difficult to ascertain; this component may be a highly differentiated rock such as rhyolite, or a moderately differentiated one such as diorite²⁹. It cannot be of typical basaltic nature because if it were the mixture with the crustal component would change the isotopic ratios so that the mixture would inherit a 'granitic' isotopic signature although it could be close to a basalt from the point of view of major element chemistry. For neodymium a crust/mantle concentration ratio of about 1 or 2 is reasonable. For strontium we can assume a crust/mantle concentration ratio

of 0.3 which corresponds to the granite/diorite concentration ratio.

Most of our data plot in the field between the young and old crust mixing lines. For every young granite in Fig. 3, we can read in Fig. 4 its specific mixing ratio m^c/m^m . For a given ordinate in Fig. 4, a conservative estimate of the crust/mantle mixing ratio is provided by the interval between the ϵ_{Nd} values for the young and old crust mixing curves. A qualitative description in terms of young or old crust recycling can be obtained, rather than a precise young/old crust ratio or crust age. However, when the crust/mantle mixing ratio is found to be close to unity, a good approximation of crustal age is given by the Sm/Nd model age.

Most orogenic granitoids have been mixed with relatively young continental crust (Sm/Nd model ages: 600–1,200 Myr); this includes the Japan, Sierra Nevada, European Hercynian granitoids and Andean granitoids or andesites. For these areas the basement is not expected to be very old and, therefore, the results agree with the regional geology. The continental/mantle component mass ratio values are probably between 1 and 4. The highest ratio is that for the Himalayan anatectic granites which represent almost pure remelting of 1,100–2,300 Myr old basement. The hercynian granites from the Pyrenees have also a high m^c/m^m ratio.

The southern California batholith, on the other hand, plots outside the upper crust mixing curve field and would be accounted for by a mixture of mantle material and a lower crustal

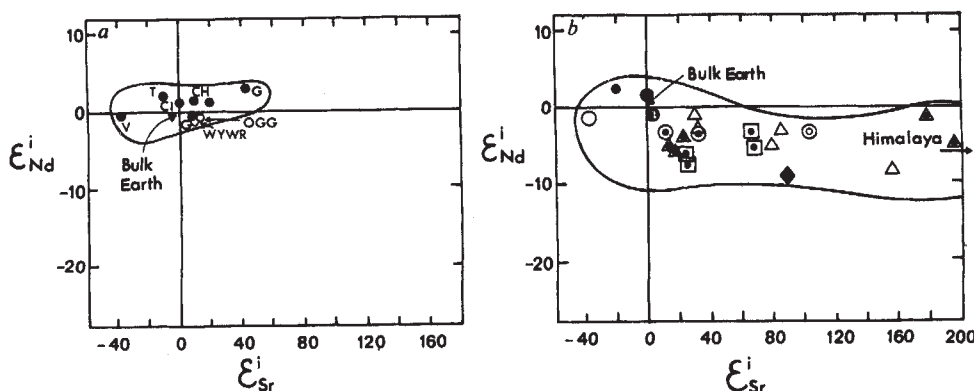


Fig. 3 a, b Comparison of ϵ_{Nd}^i with ϵ_{Sr}^i , showing the contrast in the distribution of experimental points between granitoids older than 1,000 Myr (a) and those younger than 400 Myr (b).

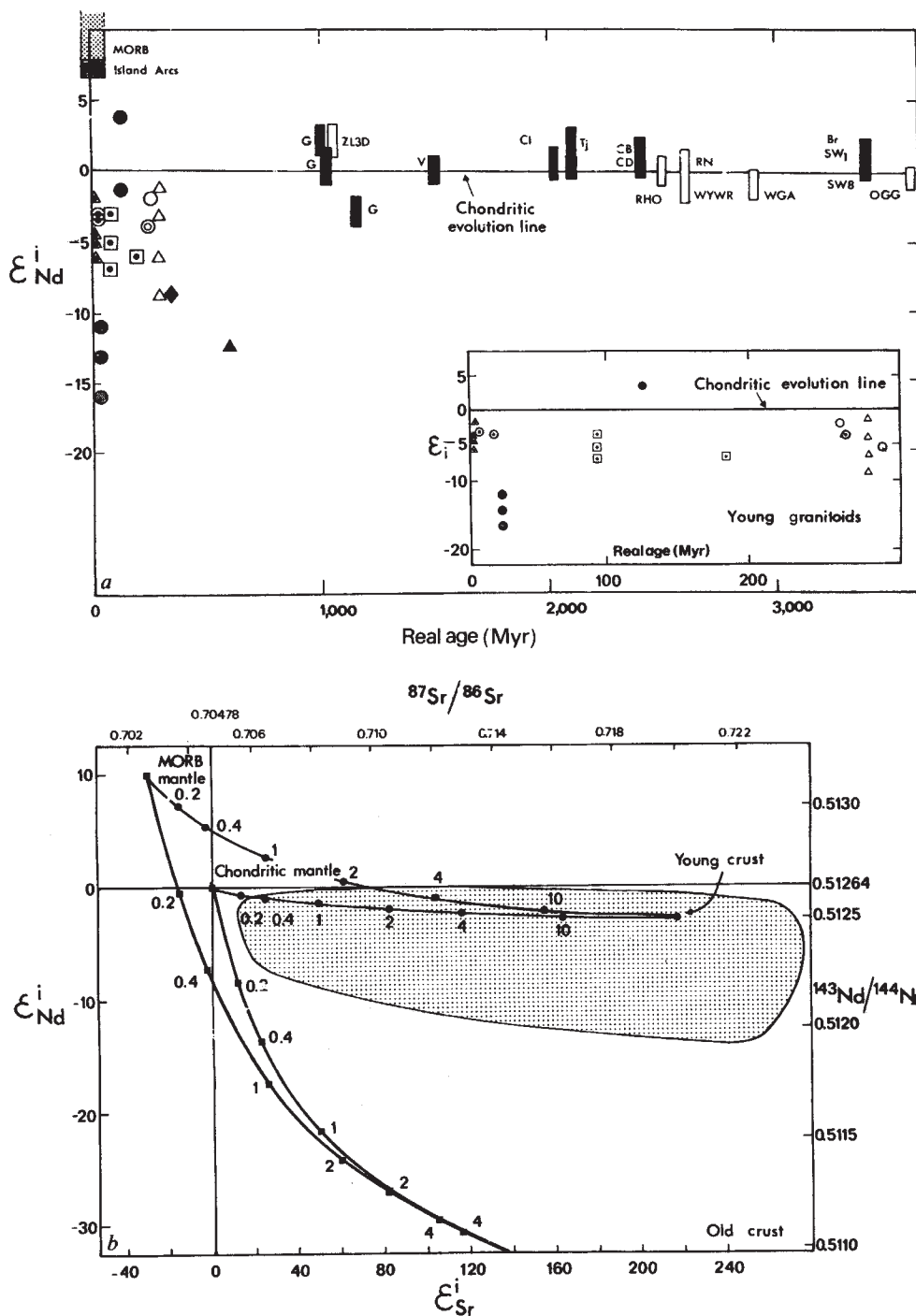


Fig. 4 a, Plot of ϵ_{Nd}^i with time (data from refs 9-12 and from the present study). As in Fig. 3 we can see the increase in deviation from the chondritic line with time. Young granites symbols are as in Fig. 1. Precambrian granites: this work: Cl, Ivory Coast; Tj, Tassendjanet; C_B, C_D, North East China; SW₁, SW₂, Swaziland; V, Venezuela; G, Grenville. Other results: ZL 3D, Llano uplift; RHO, Great Dyke; RN₃, Preissac Lacorne B; WYWR, Louis Lake Gr₁ Wyoming; WGA, Anorthosite West Greenland; OGG, Amitsoq gneisses. b, Comparison of ϵ_{Nd}^i with ϵ_{Sr}^i illustrating the mixing model. The mixing is between two possible mantle materials (one MORB, one chondritic) and two possible continental crusts (old and young). Mixing equations:

$$R_{Nd}^m x + R_{Nd}^c (1-x) = R_{Nd}^z$$

$$R_{Sr}^m y + R_{Sr}^c (1-y) = R_{Sr}^z$$

where R_{Nd} is $^{143}Nd/^{144}Nd$ and R_{Sr} is $^{87}Sr/^{86}Sr$; R^m corresponds to the mantle input, R^c to the crustal input and R^z to the mixture.

$$x = \frac{C_{Nd}^m m^m}{C_{Nd}^m m^m + C_{Nd}^c m^c}$$

$$= \frac{1}{1 + (C_{Nd}^c / C_{Nd}^m)(m^c / m^m)}$$

$$y = \frac{1}{1 + (C_{Sr}^c / C_{Sr}^m)(m^c / m^m)}$$

C_{Nd}^c (C_{Sr}^c) is the concentration of Nd(Sr) in the crustal component; C_{Nd}^m (C_{Sr}^m) the concentration of Nd(Sr) in the mantle component; m^c the mass of the crustal component; m^m the mass of the mantle component. The two curves indicate:

$$\frac{[Sr] \text{ in mantle material}}{[Sr] \text{ in continental material}} = 3.3$$

$$\frac{[Nd] \text{ in mantle material}}{[Nd] \text{ in continental material}} = 0.75$$

$$\frac{m_c}{m_m} = \frac{\text{mass of continental material}}{\text{mass of mantle material}}$$

The hatched area corresponds to the observed data.

component; such a conclusion does not contradict that of Taylor and Silver¹⁸, but a more comprehensive study is needed to really understand the data. The position of the Skye anorogenic granite in the (ϵ_{Nd}^i , ϵ_{Sr}^i) diagram is close to the field of the Lewisian gneissic basement³⁷. Its genesis has therefore involved remelting of an old basement.

The Nd and Sr isotopic results for most young granitoids leads to the main conclusion that there has been extensive involvement of previous continental crust. This effect is particularly important for continental collision orogenesis (almost pure crust remelting in the MCT from Himalayas). Subduction type orogeny involves more mantle material but with m^c/m^m values of 0.5-2. We can stress three points concerning the mixing process.

(1) Where the mixing³⁸ occurs has not been determined. It may take place in the geosyncline between volcanic rocks and detrital sediments, or during the intrusion of mantle diorite by assimilation, or even by sediments dragged down by the downgoing slab.

(2) Except for continental collision areas, the basement involved is not extremely old, which agrees with the border type models of continental growth (see below).

(3) Because Nd is extremely insoluble and not very mobile during water transport or metamorphism⁴¹ the mixing process cannot be interpreted as selective contamination as it was for Sr or Pb. We had then to consider a volume-to-volume mixture. This is a characteristic of Nd isotopes and reinforces their usefulness.

Genesis of Precambrian granites

The Sm-Nd isotopic relationship of several Precambrian granitoid rocks have been investigated. They include:

- Isua gneisses, 3,720-Myr old⁴²⁻⁴⁴.
- Amitsoq gneisses from Greenland, 3,650-Myr old⁴¹⁻⁴⁵.
- Preissac Lacorne batholith from Quebec⁴⁶ and Lake Louis batholith from Wyoming, ~2,650-Myr old⁴⁷.

Lewisian gneisses, 2,920-Myr old⁴⁸.

Mountain Pass, California, with proposed⁴⁹ ages of between 1,400 and 1,700 Myr.

Llano uplift, 1 Myr old⁵⁰.

We have added several new examples:

The Swaziland (South Africa) granites, 3,400-Myr old⁵¹.

The Tsien-Si metagranite (Republic of China), 2,400-Myr old.

The Tassendjanet granite (Sahara), 2,100-Myr old⁵².

The Baoulé granite from Ivory Coast, 2,000-Myr old⁵³.

The Grenville granites from Ontario, 1,000–1,300-Myr old⁵⁴.

The Rapakivi granite from Venezuela, 1,500-Myr old⁵⁵.

The Pan African granodiorites from Morocco and Ahaggar, 600-Myr old⁵².

For each granite, it is possible to measure the $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{147}\text{Sm}/^{144}\text{Nd}$ ratios. If the geological age T_1 is known independently, the initial ratio can be calculated:

$$(^{143}\text{Nd}/^{144}\text{Nd})_{i,T_1} = (^{143}\text{Nd}/^{144}\text{Nd})_{\text{meas}} - (^{147}\text{Sm}/^{144}\text{Nd})(e^{\lambda T_1} - 1)$$

De Paolo and Wasserburg¹⁰ have shown that this calculation is reasonably reliable for Sm–Nd because this chronometer is practically insensitive to weathering and to weak metamorphism. Therefore, only a simple *in situ* age correction is needed to calculate all initial ratios $(^{143}\text{Nd}/^{144}\text{Nd})_{i,T_1}$. A convenient way of comparing the deviations from the planetary evolution curve is to use the initial $\varepsilon_{\text{Nd}}^1$ notation¹⁰.

All of these granitoids have initial $^{143}\text{Nd}/^{144}\text{Nd}$ values which plot very close to the closed system planetary evolution line defined by De Paolo and Wasserburg¹⁰.

Contrary to what is observed for recent granitoids, Nd isotopes favour a mantle origin for the Precambrian granitic rocks. Our present statistics on these initial ratios are poor and need to be improved. However, the general result is impressive compared with the results for modern granites, especially as our sampling covers different petrographic types in the Precambrian. The first interpretation would be that these granites were directly derived from the mantle through magmatic differentiation or segregation processes. However, we could also imagine a derivation from the mantle via a sedimentary cycle. In such processes, the sediments are volcanogenic greywackes which are melted a short time after their deposition. If this is the common process for granitoid formation, then our results indicate the participation of a large proportion of mantle material in the sedimentation during the Precambrian period. In some way these granites were freshly derived from the mantle and, therefore, the continental crust corresponding to these granites can be considered to be young.

As for modern granites, one important way of examining the Precambrian origin is to use the coupled Sr–Nd isotopes. As the initial ratio is more difficult to obtain for Sr because of the mobility of Rb and Sr, a complete study including whole rock isochrons is necessary to obtain such values.

The data for Precambrian granitoids plot around the origin of the diagram (Fig. 5a), even if the resulting Sr errors are larger than the Nd ones. Such a plot seems to indicate a tendency for departure from the planetary (bulk Earth) values with time, more clearly for Sr than for Nd. Comparing these diagrams with our mixing model reveals large differences in the isotopic geochemistry of Sr and Nd, and clarifies the scatter around the origin through geological time. The lower crust (granulite facies) has positive ε_{Sr} , but quite negative $\varepsilon_{\text{Nd}}^0$ (ref. 37). The fact that all experimental points plot around the origin of the Nd diagram eliminates the possibility of Precambrian granitoid originating from recycling of the lower crust⁵⁷.

The Nd and Sr evolution curves for the oceanic mantle have been determined by Zindler *et al.*⁵⁸ and C.J.A. *et al.*⁵⁹. It is very interesting to plot on the same diagram the oceanic mantle evolution curve and the evolution curve for the initial ratio of granitoids (continental crust, see Fig. 4a). For the past 2,000 Myr, the younger the age, the more the oceanic mantle and continental crust curves diverge; this means that granitoids

are not the result of pure oceanic mantle differentiation, thus recycling has been involved. This strongly suggests that the oceanic mantle isotopic evolution is due to the extraction of some material which is stored in the continental crust. A second observation is that the scatter of values for granitoids increases with time, so that the process is a statistical one. Third, we do not observe a sharp transition between old and young granitoids, but rather a continuous one. Thus, 1,000 or 500-Myr old granitoids are very important for studying the onset of the present divergence of continental crust and mantle evolution.

Various models^{3,4,57,60}, in which old continental crust was differentiated and then recycled through geological time, are ruled out by Nd data. Other models², which suggest that continental crust is derived from the mantle by a differentiation process, are not appropriate either (see Fig. 4b). The model of Moorbath⁷ is difficult to reconcile with Nd results on young granitoids. Nd contamination is not selective or accidental, as it can be for Sr, and large masses of crustal component are involved in the genesis of young andesites or granodiorites.

One interpretation of our observations is by postulating two types of granite production processes: (1) a Precambrian type for which granitoids are derived from the mantle through complex magmatic fractionation processes (calc-alkaline differentiation); (2) a post 1,000 Myr type during which crustal anatexis became the major process.

Such a sharp transition is difficult to reconcile with petrographic observations since Moorbath *et al.* (ref. 61 and unpublished results) have clearly demonstrated the existence of crustal anatexis in the Archean, and others^{62–64} have clearly demonstrated, for various types of granitoids linked with a calc-alkaline suite, that many Precambrian granites have close equivalents in modern ones.

In addition, such a sharp transition will bring back the 'catastrophism' problem usually poorly supported by geological observations. We will, therefore, assume a uniform set of processes for the genesis of granitoids through geological times, with an evolution in the characteristics of these processes. We suppose that the variation applies to the composition of the materials from which the granitoids derive.

Mantle–crust mixing model for granitoid formation and continental growth

We have seen that isotopic data for young granitoids can be understood through a mixing model involving a mantle component and a recycled crust component. Can this model be extended to the Precambrian granitoids?

Adopting the (*I*–*S*) classification of White and Chappell⁶⁵ which agrees with Didier and Lameyre's classification based on xenoliths⁶⁶, we examine the genesis of these two types of granitoids.

The (*S*) granitoids are derived from sediments or metasedimentary rocks. All sediments are composed of a recycled detrital component and a mantle component. For sediments which are abundant in metasedimentary piles such as greywackes or schists this type of dual origin is clear. In the Archean, the continental crust was not as developed as it is today: the mantle component was dominant in such sediments. With time, however, the isotopic signature shifted from a mantle value to a continental crust one. In recent times the continental component has dominated. Such a model explains quite well the observations by S. R. Taylor on shales. These authors claim that these are pure continental crust sediments while we believe that they are mixtures with a mantle component⁶⁷.

We admit that the (*I*) granitoids result from a mixture between an intrusive component of mantle origin (diorite) and its surroundings. These surroundings are composed of sediments or metasediments or granitoids.

The above reasoning may then be applied, but with a larger mantle contribution. Thus, both (*I*) and (*S*) granitoids can be explained through a mixing model, but does our model fit the data quantitatively?

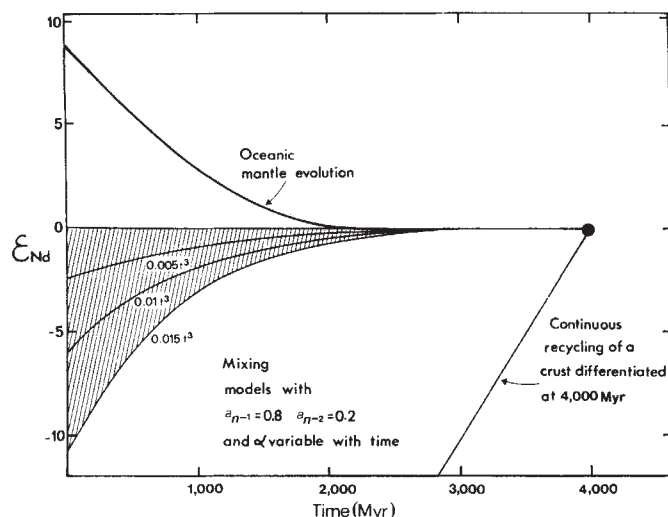


Fig. 5 Mixing model curves. Computed model with two last orogeneses entering the continental crust (last orogeny contribution is four times that of the one before). The degree of mixing between mantle and preexisting crust is assumed to be increasing with time as t^3 , with more and more crust entering the orogenesis. The curves for oceanic mantle and continuously recycled crust differentiated at 4,000 Myr are also plotted as references.

Our hypothesis is based on the fact that continents have grown through geological time, although not always with the same rate. This growth takes place during orogeneses. These orogeneses are discontinuous events, at least locally. Therefore, we will use a finite difference method. Let $R_n(t)$ be the isotopic ratio (Nd or Sr) at time t of the granitoid formed during the n th orogenesis which occurred at time t_n (or age t_n). Let $R_{c,n}(t)$ be the isotopic ratio at time t (or age t) of the mantle. If α is the mass fraction of Nd (or Sr) from the preexisting crust involved in the mixing, then

$$R_n(t_n) = \alpha R_{c,n}(t_n) + (1 - \alpha) R_m(t_n)$$

Because geological observations^{68,69} and observations made on California and Laddakh granitoids suggest that the oceanic mantle is involved in continental growth, we will take the oceanic mantle isotopic ratio for $R_m(t)$. The crustal composition will be written as a polynomial:

$$R_{c,n}(t_n) = \sum_{i=0}^{i=n-1} a_i (R_i(t_i) + \mu_c \lambda (t_n - t_i))$$

with

$$a_i > 0 \quad \text{and} \quad \sum_i a_i = 1$$

and

$$\mu_c = \frac{^{147}\text{Sm}}{^{144}\text{Nd}}$$

in the continental crust. $a_i(R_{c,i}(t_i) + \mu_c \lambda (t_n - t_i))$ is the crustal contribution of the i th orogenesis to crustal composition. This approach uses the fact that μ_c is almost constant in granitoids of different ages.

The parameters a_i can be chosen in different ways: if we take $a_{n-1} = 0.8$, $a_{n-2} = 0.19$, $\sum_{i=1}^{n-3} a_i = 0.01$, recycling is restricted to the last two orogeneses. This hypothesis is supported by the erosion being proportional to the height of mountains, and is therefore only important for young mountains. It also agrees with the theory of continental growth by successive belts. We can choose a smoother law:

$$a_{n-1} = 0.4, \quad a_{n-2} = 0.3, \quad a_{n-3} = 0.2, \quad a_{n-2} = 0.08 \dots$$

This assumption is supported by the fact that sediments are created by large rivers such as the Mississippi, the Yang Tse, the Mekong, and the Amazon, which flow through terrains of various ages.

Another hypothesis about the mixing parameter must be made. Although we assume that α is constant through geological time (this is the steady state approximation), it is more likely to have increased because during the Precambrian the mantle was more active⁷⁰⁻⁷² and the continental crust not so developed and thus there were more volcanogenic sediments than at present. Observations of Archean sediments support this point of view⁷⁰.

As we have no evidence that continental differentiation started before 4,000 Myr, our calculations start from that date. Orogeneses are supposed to occur with a regular time interval of 500 Myr. In fact, each parameter is susceptible to variations and our model should be stochastic. However, the model is given here in its mean value form. Numerical calculations of these equations and comparison with observed data have shown that: (1) a mixing ratio α , which is constant through geological times does not fit the data; (2) the function α must depend on time as t^3 or t^4 , but we cannot be more specific with the available data. (3) With the present data we cannot tell whether the last two or four orogeneses are involved in crust formation. The involvement of the four last crustal segments in the orogeneses would imply that the mixing coefficient is lower than if only two are involved. (4) The polynomial coefficient of an old orogeny may possibly become more important with time. This could occur if the Precambrian orogeny were to follow the centrifugal growing and then in modern time the plate tectonics focate the orogeny in random way relative to the age of crustal segments.

Figure 5 shows ϵ_{Nd} variations with time for $a_{n-1} = 0.8$ and $a_{n-2} = 0.2$ and different functions α . Hence with certain conditions the mixing model is seen to apply and now requires a more refined study, especially in the Precambrian.

Conclusion

Our model quantitatively fits the isotopic variation with geological age and explains the fact that Pb isotopes are a record of continental recycling whereas Sr is more appropriate for mantle derivation, Nd being an intermediate tracer.

We thank H. P. Taylor, P. M. Hurley, R. Kistler and Z. Peterman, J. R. Lancelot, P. Lefort, P. Brunel and S. Moorbath for samples, also P. M. Hurley, S. R. Hart, S. Moorbath, J. F. Minster, F. Albarede, and S. Fourcade for helpful discussions. D. Rousseau made some of the Sr isotope measurements. We thank P. Richard for his help in applying the Nd technique and O. Brevart for programming. S. Fourcade, O. Brevart and J. F. Minster reviewed the manuscript. This is contribution IPG 414.

Received 2 February; accepted 23 May 1980.

- Engel, A. E. *J. Science* **140**, 143-152 (1963).
- Hurley, P. M., Huges, H., Faure, G., Fairbairn, H. W. & Pinson, W. H. *J. geophys. Res.* **67**, 5315-5334 (1962).
- Faure, G. & Powell, J. L. in *Strontium Isotope Geology*, 188 (Springer, Berlin, 1972).
- Patterson, C. C. in *Isotopic and Cosmic Chemistry* (ed. Wasserburg, C. M.) 244 (North-Holland, Amsterdam, 1964).
- Allègre, C. J. in *Structure et Dynamique de la Lithosphère* (eds Allègre, C. J. & Mattauer, M.) 372-433 Hermann, Paris, 1972.
- Armstrong, R. L. *Rev. Geophys.* **6**, 175-199 (1968).
- Moorbath, S. *Chem. Geol.* **20**, 151-187 (1977).
- Richard, P., Shimizu, N. & Allègre, C. J. *Earth planet. Sci. Lett.* **31**, 269-278 (1976).
- De Paolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 743-746 (1976).
- De Paolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249-252 (1976).
- O'Nions, R. K., Hamilton, P. J. & Evensen, N. H. *Earth planet. Sci. Lett.* **34**, 13-22 (1977).
- De Paolo, D. J. & Wasserburg, G. J. *Geol. Res. Lett.* **4**, 465-468 (1977).
- Hawkesworth, C. J. *et al. Earth planet. Sci. Lett.* **36**, 253-262 (1977).
- Silver, L. T. *Geol. Soc. Am. Abstr.* **13**, 193-194 (1977).
- Silver, L. T., Allen, C. R. & Stehle, F. G. *Geol. Soc. Am. A. Meet.* 179-280 (1968).
- Silver, L. T. & Early, T. O. *EOS* **58**, 532 (1977).
- Silver, L. T., Early, T. O. & Anderson, T. H. *Geol. Soc. Am. Abstr.* **7**, 375-376 (1975).
- Taylor, H. P. & Silver, L. T. *Int. Conf. Geochron. Cosmochron. Isotope Geol.* Aspen, 423-426 (1978).
- Hurley, P. M., Bateman, P. C., Fairbairn, H. W. & Pinson, W. H. *Jr Bull. geol. Soc. Am.* **76**, 165-174 (1965).
- Kistler, R. W., Peterman, Z. E., Ross, D. C. & Gothfried, D. *Geological Science XIII* (School of Earth Science, Stanford University, 1973).
- Magaritz, M. & Taylor, H. P. *Jr Can. J. Earth. Sci.* 1514-1536 (1976).
- Armstrong, R. D., Taubeneck, W. H. & Hales, P. Q. *Bull. geol. Soc. Am.* **88**, 397-411 (1977).
- Lancelot, J. R., Briqueu, L., Westphal, B. & Tatsumoto, M. *Int. Conf. Geochron. Cosmochron. Isotope Geol.*, Aspen, 240-241 (1978).
- Briqueu, L. & Lancelot, J. R. *Earth planet. Sci. Lett.* **43**, 385-396 (1979).
- Mattauer, M. *Earth planet. Sci. Lett.* **28**, 144-154 (1975).
- Lefort, P. *Rev. Geogr. Phys. Geo. Dyn.* **13**, 5-12 (1971).
- Hamet, J. & Allègre, C. J. *Bull. geol. Soc. Am.* **84**, 1429-1442 (1976).

28. Andrieux, J., Brunel, M. & Hamet, J. *Colloq. int. Cent. natn. Rech. Scient.* **268**, 31–40 (1976).
29. Vitrac-Michard, A. & Allègre, C. J. *Contr. Miner. Petrol.* **50**, 257–285 (1975).
30. Vitrac-Michard, A. & Allègre, C. J. *C. r. hebdom. Séanc. Acad. Sci., Paris* **269**, 2174–2177 (1969).
31. Autran, A., Fonteuilles, M. & Guitard, G. *Bull. Soc. fr. Géol.* **XII**, 4, 673–731 (1970).
32. Le Terrier, J. thesis, Univ. Nancy (1972).
33. Lacroix, A. *Bull. Serv. Cart. Géol. Fr.* **II**, 71, 501–550 LXVII (1900).
34. Mattauer, M. *Mémoire Réunion. Ann. Sci. Terre*, Nancy, 279 (1974).
35. Fourcade, S. & Allègre, C. J. (in preparation).
36. Allègre, C. J., Ben Othman, D., Polvé, M. & Richard, P. in *Phys. Earth planet. Inter.* **19**, 293–306 (1979).
37. Carter, S. R., Evensen, N. M., Hamilton, R. K. & O'Nions, R. K. *Science* **202**, 743–747 (1978).
38. Allègre, C. J. & Dars, R. *Geol. Rdsh.* **55**, 226–237 (1965).
39. Heier, K. S. *Fortsch. Miner.* **50**, 174–187 (1974).
40. Dupuy, C., Leyreloup, A. & Vernières, J. *2ème Symp. Origin and Distribution of Trace Elements* (Unesco, Paris, 1977).
41. Haskin, L. A. et al. *J. geophys. Res.* **71**, 6091–6105 (1966).
42. Moorbath, S., O'Nions, R. K. & Pankhurst, R. J. *Earth Planet. Sci. Lett.* **27**, 229–239 (1975).
43. Hamilton, P. J., O'Nions, R. K. & Evensen, N. H. *Nature* **272**, 41 (1978).
44. Vitrac-Michard, A., Lancelot, J. R., Allègre, C. J. & Moorbath, S. *Earth planet. Sci. Lett.* **35**, 449–453 (1977).
45. Moorbath, S. & Pankhurst, R. J. *Nature* **262**, 124–126 (1976).
46. Steiger, R. H. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **33**, 1212–1232 (1969).
47. Naylor, N. S., Steiger, R. H. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **34**, 1133–1154 (1970).
48. Hamilton, P. J., Evensen, N. H., O'Nions, R. K. & Tarney, K. *Nature* **277**, 25–28 (1979).
49. Lamphere, M. A. *J. Geol.* **72**, 381–399 (1964).
50. Zartmann, R. E. *J. Petrol.* **5**, 359–408 (1968).
51. Allsopp, H. L., Ulrych, T. J. & Nicolaysen, L. O. *Can. J. Earth Sci.* **5**, 605 (1968).
52. Allègre, C. J. & Caby, C. R. *C. r. hebdom. Séanc. Acad. Sci. Paris* **275**, 2095–2098 (1972).
53. Bonhomme, thesis, Univ. Clermont Ferrand (1961).
54. Krogh, T. E. & Hurley, P. M. *J. geophys. Res.* **73**, 7107–7118 (1968).
55. Gaudette, H. E., Mendoza, V., Hurley, P. M. & Fairbairn, H. W. *Bull. geol. Soc. Am.* **89**, 1335–1340 (1978).
56. Hargraves, R. B. *Science* **143**, W 4251 363 (1976).
57. O'Nions, R. K. & Pankhurst, R. J. *Earth planet. Sci. Lett.* **38**, 211–236 (1978).
58. Zindler, A., Brooks, C., Arndt, T. N. & Hart, S. *Int. Conf. Geochron. Cosmochron. Isotope Geol. Aspen*, 469–471 (1979).
59. Allègre, C. J., Brevard, O., Dupré, B. & Minster, J. F. *Phil. Trans. R. Soc.* (in the press).
60. Fyfe, W. S. *Geol. Sci. Can.* **3**, 82–83 (1976).
61. Moorbath, S., Wilson, J. F. & Cotterill, P. *Nature* **264**, 536–538 (1976).
62. O'Nions, R. K. & Pankhurst, R. J. *Earth planet. Sci. Lett.* **22**, 328–338 (1974).
63. Tarney, J., Dalziel, I. W. & De Wit, M. J. *The Early History of the Earth* (ed. Windley, B. F.) 131–146 (London, 1976).
64. Barker, F., Wones, D. R., Sharp, W. N. & Desborough, G. A. *Precamb. Res.* **2**, 97–160 (1975).
65. White, A. J. R. & Chappell, R. W. *Tectonophysics* **43**, 7–22 (1977).
66. Didier, J. & Lameyre, J. *Contr. Miner. Petrol.* **24**, 219–238 (1969).
67. Taylor, S. R. *Chemical Composition and Evolution of the Continental Crust in The Earth: its Origin, Structure and Evolution* (ed. McElhinny, M. W.), 353–376 (Academic, New York, 1979).
68. Michot, P. *Revue universelle Mines* **9**, 249–261 (1968).
69. Goodwin, A. *Can. J. Earth Sci.* **12**, 2737–2759 (1977).
70. Veizer, J. & Jansen, S. *J. Geol.* **87**, 341–510 (1979).
71. Taylor, S. R. *Tectonophysics* **4**, 17–34 (1967).
72. Brevard, O. & Allègre, C. J. *Geol. Soc. Fr.* **16** (1979).
73. Krogh, T. E. *Yb. Carnegie Instn Wash.* **69**, 342–344 (1971).
74. Lugmair, G. W., Scheimin, N. B. & Marti, K. *Earth planet. Sci. Lett.* **27**, 79–84 (1975); *Proc. 6th Conf. Lunar Sci.* 1419–1429 (1975); *Proc. 7th Conf. Lunar Sci.* 2009–2033 (1976).
75. Birc, J. L. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 37–51 (1978).
76. Vitrac-Michard, A. thesis, Univ. Paris (1975).

Continental crust composition and nature of the lower crust: constraints from mantle Nd–Sr isotope correlation

Barry L. Weaver & John Tarney

Department of Geology, The University of Leicester, Leicester LE1 7RH, UK

Calculations suggest that if the widespread upper mantle depletion in lithophile elements, as represented by the negative correlation between Nd and Sr isotopes in basalts, has resulted from continental crustal extraction, then extraction of observed Archaean crustal compositions best explains the Nd–Sr isotope relationship. Much of the lower continental crust must, therefore, have the high Nd/Sm and low Rb/Sr ratios typical of Archaean granulites.

RECENT developments in neodymium isotope geochemistry^{1–3}, and the combined use of both neodymium and strontium isotope data^{4–9} on mantle-derived igneous rocks, have produced information on the Rb/Sr and Sm/Nd history of their mantle source regions^{6–9}. Clearly much of the present upper mantle has suffered a time-integrated depletion of Rb relative to Sr and of Nd relative to Sm. Corresponding evidence on Archaean igneous rocks^{2,4,10–12} suggests, however, that the Archaean mantle had broadly chondritic Sm/Nd ratios. Because the continental crust represents a reservoir which is enriched in Rb relative to Sr and in Nd relative to Sm and, overall, has higher ⁸⁷Sr/⁸⁶Sr and lower ¹⁴³Nd/¹⁴⁴Nd ratios than the mantle, and because there is strong isotopic evidence that the crust has grown with time since ~4,000 Myr ago^{13–15}, it is a reasonable assumption that extraction of continental crust may have caused the observed Rb/Sr and Nd/Sm depletion relative to primordial ratios in the present mantle. There is no unambiguous evidence that another major high Rb/Sr and high Nd/Sm reservoir exists which might complement the generally depleted nature of present upper mantle sources. Preliminary attempts have therefore been made to assess the effects of crustal extraction¹⁶ (and crustal contamination¹⁷) on the Sr and Nd isotope systematics of the upper mantle reservoir.

Here we examine the crust–mantle isotopic relationship in more detail, but with respect to specific Precambrian crustal materials rather than quoted 'average' crustal compositions. This approach is more relevant because (1) the time scale on which mantle can be completely recycled is probably of the order of several hundred million years, (2) considerable crustal growth occurred during the early Precambrian^{14,15}, (3) the long

half lives of ⁸⁷Rb and ¹⁴⁷Sm necessitate a long time lapse before changes in ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd ratios of the magnitude observed in the present-day mantle can develop¹⁸, and (4) there are significant differences between Archaean and Phanerozoic crustal compositions¹⁹.

The crustal data used in the calculations are based on the Archaean Lewisian complex of north-west Scotland, a region which contains juxtaposed segments of presumed upper and lower crustal orthogneisses²⁰, the geochemical^{20–24} and isotopic (Sr, Nd and Pb)^{12,25–30} characteristics of which are relatively well known.

Sr–Nd isotope correlation and crustal extraction

Measured Sr and Nd isotopic ratios in many young basalts exhibit a strong negative correlation^{5,6} shown by the 'main trend' (defined by $\epsilon_{Nd} = -2.7 \epsilon_{Sr}$) (ref. 16; see also Fig. 3) where ϵ_{Nd} expresses the deviation (in parts per 10⁴) of ¹⁴³Nd/¹⁴⁴Nd ratios observed in igneous rocks at present from the ¹⁴³Nd/¹⁴⁴Nd ratio expected in a reservoir which has always had chondritic Sm/Nd (ref. 16). ϵ_{Sr} is defined similarly for ⁸⁷Sr/⁸⁶Sr ratios relative to a reservoir with bulk Earth Rb/Sr ratios¹⁶. Thus rocks derived from a mantle source which has suffered a time-integrated depletion of Rb relative to Sr and Nd relative to Sm will plot in the + ϵ_{Nd} , – ϵ_{Sr} quadrant on the ϵ_{Nd} – ϵ_{Sr} diagram. Mid-ocean ridge basalts invariably plot in this quadrant, as do many ocean island and continental basalts, although these latter tend to plot nearer to the bulk Earth composition. Basalts derived from mantle sources which seem to have been altered by

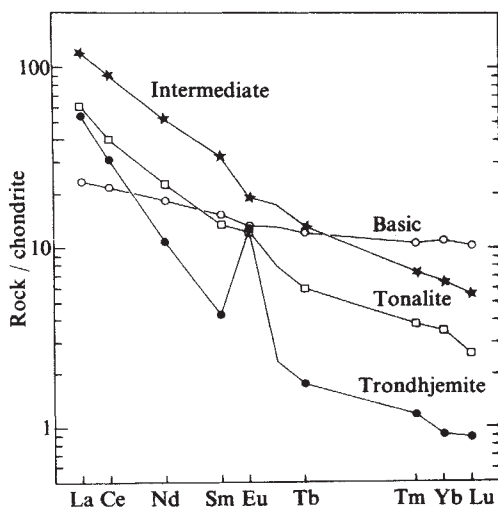


Fig. 1 Average rare-earth element patterns of the main rock types comprising the Lewisian lower crustal granulite complex, north-west Scotland (data from refs 12, 20–24, 37).

more radiogenic seawater Sr (ref. 31), or by continental material^{17,32} are displaced from the 'main trend'.

The widespread upper mantle depletion in incompatible elements represented by the 'main trend' must be accommodated in models of mantle evolution. DePaolo and Wasserburg¹⁷ consider the slope of the 'main trend' to be incompatible with upper crustal contamination of a uniform mantle reservoir, while Allègre *et al.*¹⁸ have demonstrated that the mantle main trend correlation reflects melt extraction from the mantle rather than, for instance, mixing processes involving separate mantle reservoirs. In this case the slope of the Sr–Nd isotope correlation constrains the Rb/Sr and Sm/Nd ratios of the material extracted provided an assessment can be made of the Sr/Nd ratios of the extract and of the initial mantle source¹⁶. By assuming that the continental crust is the complement to the 'main trend' mantle depletion, and by noting that there was only very limited variation in the Sm/Nd and Sr/Nd ratios of the various published estimates of 'average' upper crust and total crust composition, DePaolo¹⁶ was able to use the Sr–Nd isotope correlation to calculate the mean Rb/Sr ratio of the continental crust as lying within the range 0.044–0.10. As upper crustal compositions are quite well defined³³, it is the composition of the lower crust which is crucial in constraining models for the generation of the continental crust.

Crustal compositions

The Lewisian granulite facies gneiss complex is a prime candidate for Archaean lower crust, having the appropriate seismic characteristics^{34,35} and having equilibrated in temperature–pressure conditions as high as 1,000 °C and 15 kbar (ref. 36), corresponding to depths of the order of 45 km. Nd isotope studies¹² have demonstrated that the immediate igneous precursors of these gneisses were generated from a mantle source with chondritic Sm/Nd ratios 2,920 Myr ago.

The complex is a heterogeneous mixture of ultramafic, mafic, intermediate, tonalitic and trondhjemitic gneisses, but tonalitic–trondhjemitic compositions dominate. Average rare-earth element (REE) patterns³⁷ for the main rock types are shown in Fig. 1. The tonalitic and trondhjemite patterns in particular are highly fractionated, with high Ce_N/Yb_N ratios (and Nd/Sm ratios) owing to the strong depletion in the heavy REE caused by residual garnet in the source during the partial melting event which produced the gneiss precursors^{24,37}. Using these REE data plus previously published results for La, Ce, and Y (determined by X-ray fluorescence) on several hundred granulite

samples²⁰, it is possible to calculate a weighted mean REE pattern for Lewisian lower crustal granulites (Fig. 2). Because of their very high Nd/Sm ratios and age, the Lewisian granulites are markedly unradiogenic with respect to Nd isotope ratios¹² in comparison with the bulk Earth. The granulites are also high in Sr, but are unusually low in Rb and other radioactive heat producing elements (K, U, Th) largely as a result of the removal of these elements by an active fluid phase during the granulite facies event^{12,38}. Thus their Rb/Sr ratios are unusually low, and indeed the present day ⁸⁷Sr/⁸⁶Sr ratios of many granulites are below typical mantle values³⁹.

The amphibolite-facies gneiss sections of the Lewisian complex have a much smaller component of ultramafic, mafic and intermediate gneisses, and their mean major- and trace-element compositions are essentially identical to the tonalitic–trondhjemitic components of the granulite facies complex except that they have much higher Rb contents and Rb/Sr ratios. Although full representative REE data are not yet available for these gneisses, comprehensive X-ray fluorescence data for La, Ce and Y (refs 20–24), and limited data for Sm and Nd (ref. 12), allow a fairly accurate estimation of their Sm/Nd ratios, which must be close to the values for the tonalitic–trondhjemitic granulites. From their tectonic and metamorphic state we would regard these gneisses as being more representative of an intermediate crustal level rather than upper crust.

For comparison with these compositions we have also used the mean 'upper crust' composition and the corresponding 'lower crust' composition derived by Taylor³³ on the basis of the

Table 1 Continental crust compositions*

	Andesite model			Archaean	
	UC	TC	LC	LGr	LAm
SiO ₂	66.0	58.0	54.0	61.2	67.3
TiO ₂	0.6	0.8	0.9	0.5	0.3
Al ₂ O ₃	16.0	18.0	19.0	15.6	15.9
FeO*	4.5	7.5	9.0	5.3	3.4
MgO	2.3	3.5	4.1	3.4	1.2
CaO	3.5	7.5	9.5	5.6	3.4
Na ₂ O	3.8	3.5	3.4	4.4	4.6
K ₂ O	3.3	1.5	0.6	1.0	2.2
Cr	35	55	65	88	27
Ni	15	20	23	58	16
Zr	240	100	30	202	199
Rb	110	50	20	12	62
Sr	350	400	425	569	541
Ba	700	350	175	757	838
Pb	15	7	3	13	18
Th	10.5	2.5	—	0.42	9
U	2.5	1	0.25	0.05	—
La	38	19	9.5	22	29
Ce	80	38	17	44	64
Pr	8.9	4.3	2.0	—	—
Nd	32	16	8	18.5	(25)
Sm	5.6	3.7	2.75	3.3	(4.3)
Eu	1.1	1.1	1.1	1.18	—
Gd	4.7	4.2	3.95	—	—
Tb	0.77	0.64	0.58	0.43	—
Dy	4.4	3.7	3.35	—	—
Ho	1.0	0.82	0.73	—	—
Er	2.9	2.3	2.0	—	—
Tm	0.5	0.4	0.35	0.19	—
Yb	2.8	2.2	1.9	1.2	—
Lu	0.4	0.3	0.25	0.17	—
Y	27	22	20	9	9
Ce _N /Yb _N	7.3	4.4	2.3	9.3	—

UC, TC, and LC are upper crust, total crust and (calculated) lower crust compositions respectively, based on andesite model for crustal growth³³. LGr is mean composition of Lewisian granulites (= Archaean lower crust); LAm is mean composition of amphibolite-facies gneisses from Scotland and Greenland (= Archaean intermediate crustal level). Figures in brackets are interpolated values.

* Oxides in wt%, trace elements in p.p.m.

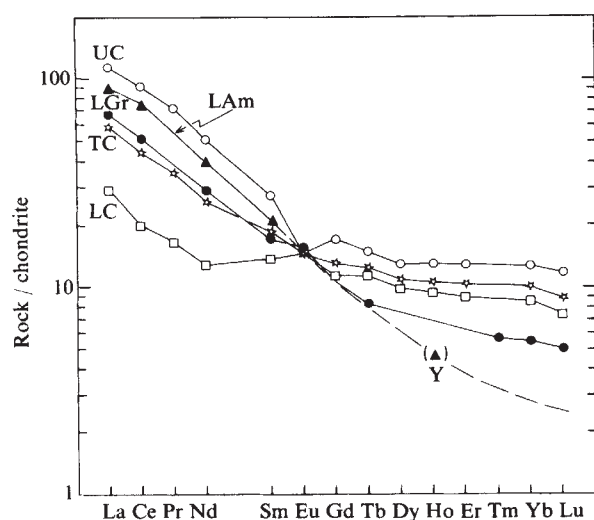


Fig. 2 Mean rare-earth element patterns of crustal components. UC, TC and LC are average upper crust, total crust and (calculated) lower crust compositions based on the andesite model for crustal growth³³. LGr is mean rare-earth pattern for Lewisian lower crustal granulites (based on Fig. 1 plus data in refs 12, 20–24, 37). LAm, mean rare-earth pattern for Lewisian amphibolite-facies gneisses (based on data in refs 12, 20, 22, 38).

andesite model for crustal growth. Relevant data for these compositions and for the Lewisian granulite and amphibolite facies gneisses are given in Table 1, and the corresponding REE patterns are plotted on Fig. 2.

Results

The modelling of $\epsilon_{Nd}-\epsilon_{Sr}$ (Fig. 3) follows the method used by DePaolo¹⁶, except that the data in Table 1 are used to derive vectors on the $\epsilon_{Nd}-\epsilon_{Sr}$ diagram for residual mantle compositions after extracting the various crustal components. Clearly a valid model should yield mantle vectors for the extraction of upper crust and lower crust which embrace the mantle 'main trend'. The vectors calculated are very dependent on the Sr/Nd ratio of mantle + crust, and less dependent on the parameter X_c (the fraction of Nd atoms in the crust + depleted-mantle system residing in the crust¹⁶). The values used here for the Sr/Nd ratio of mantle + crust and X_c are those of DePaolo¹⁶. Where upper crust and lower crust compositions have been treated separately the X_c factor has been adjusted accordingly. The method for calculating the vectors is given in Table 2 (see also ref. 16).

The slopes of the vectors for crust and mantle compositions plotted on Fig. 3 are directly proportional to the factors R_f in Table 2. The vectors for depleted mantle compositions in Fig. 3, derived respectively from the given crustal compositions, illustrate several interesting features. That resulting from the extraction of Lewisian granulites has a very steep trend. This arises because the mean Rb/Sr ratio of Lewisian granulites (0.021) is actually slightly lower than that estimated for the uniform mantle reservoir (0.029, ref. 16), hence resulting in little fractionation of ϵ_{Sr} in the residual mantle. Lewisian amphibolite facies gneisses, although having a rather higher mean Nd/Sm ratio than the granulites, have a much higher mean Rb/Sr ratio (0.13, Table 1) resulting in a much shallower slope of the residual vector.

The average upper crust composition of Taylor³³ has a similar Nd/Sm ratio to that of Lewisian gneisses, but has a high Rb/Sr ratio (0.31, Table 1), thus the corresponding residual mantle vector has a shallower slope than the mantle 'main trend' on Fig. 3. The calculated lower crust composition based on the andesite model actually plots in the $+\epsilon_{Nd}$, $+\epsilon_{Sr}$ quadrant, and the corresponding residual mantle vector has an even shallower slope than that produced by extraction of average upper crust. The andesitic lower crust, in fact, has a calculated Nd/Sm ratio less than chondritic, although the calculated full rare-earth pattern is slightly light-REE enriched (Fig. 2). However, if the upper crust was taken to be 30% of the total crust (rather than 33%), the calculated lower crust would then have a chondritic Sm/Nd ratio.

Figure 3 shows that extraction of an andesite bulk crust³³ from a mantle with chondritic Sr/Nd ratios could not produce the mantle 'main trend'. Because the vectors for the average upper crust of Taylor³³ and for the amphibolite facies Lewisian are similar, the determining factor in whether crustal extraction can reproduce the mantle Sr–Nd isotope trend is the nature of the lower crustal component. Apparently, the lower crust component in any crustal model is required to have a moderately high Nd/Sm ratio together with a very low Rb/Sr ratio—a chemical characteristic typical of Lewisian granulites. Extraction of continental crust with this composition as a component is compatible with a simple model for the generation of the mantle isotope correlation.

An independent test of mean crustal compositions can be made using the composition of the source of mid-ocean ridge basalts (MORB), assuming this to be the mantle residue complementary to the crustal extract. Thus if the Sr/Nd ratio of the MORB source can be defined it is possible to estimate the slope, on the $\epsilon_{Nd}-\epsilon_{Sr}$ diagram, of the required crystal extract. Using the data of Sun and Nesbitt⁴¹ for average MORB, together with reasonable assumptions of the residual mantle mineralogy, the degree of melting necessary to generate MORB

Table 2 Compositions of crustal components and calculated parameters used in $\epsilon_{Nd}-\epsilon_{Sr}$ modelling

	Rb	Sr	Nd	Sm	Rb/Sr	Nd _N /Sm _N	Sr/Nd	$f_{Rb/Sr}^*$	$f_{Sm/Nd}^*$	$(R_f)_c^*$	$(R_f)_{dm}^\dagger$
Upper crust ³³	110	350	32	5.6	0.31	1.82	10.9	9.8	-0.45	-0.046	-0.093
Total crust ³³	50	400	16	3.7	0.13	1.37	25.0	3.3	-0.28	-0.082	-0.038
Lower crust [‡]	20	425	8	2.75	0.047	0.92	53.1	0.62	-0.086	0.14	0.029
Lewisian amphibolite-facies gneiss§	74	580	25	4.1	0.13	1.92	23.2	3.4	-0.48	-0.14	-0.102
Lewisian granulite-facies gneiss§	12	570	19	3.3	0.021	1.82	30.0	-0.27	-0.45	1.67	0.78
Chondrite values	2.5	11	0.60	0.19	0.23	1.00	18.3	6.8	0.0	—	—

$$^* f_{Rb/Sr} = \frac{(Rb/Sr)}{0.029} - 1 \text{ and } f_{Sm/Nd} = \frac{(Sm/Nd)}{0.317} - 1; (R_f)_c = \frac{f_{Sm/Nd}}{f_{Rb/Sr}}$$

$^\dagger (R_f)_{dm}$ (dm, depleted mantle) calculated from (see ref. 16):

$$\frac{(R_f)_c}{(R_f)_{dm}} = \frac{(Sr/Nd)_c}{(Sr/Nd)_{dm}} \text{ where } (Sr/Nd)_{dm} \text{ derived from } X_c \frac{(Sr/Nd)_c}{(Sr/Nd)_{Eth}} + (1 - X_c) \frac{(Sr/Nd)_{dm}}{(Sr/Nd)_{Eth}} = 1$$

$$X_c = 0.5 \text{ and } (Sr/Nd)_{Eth} = 18.3$$

‡ Lower crust calculated from data in ref. 33 assuming upper crust and lower crust respectively one-third and two-thirds of total crust.

§ Mean Sm and Nd values calculated from data in refs 20 and 37, weighted by rock type. Note, amphibolite facies average is different to that of Table 1.

and utilizing published partition coefficient data for Rb, Sr, Nd and Sm (ref. 42), the composition of the MORB source can be calculated (Table 3). The R_t value for this source (Table 3), defining the slope on the $\epsilon_{Nd}-\epsilon_{Sr}$ diagram, is in excellent agreement with the value of -0.25 obtained for the mantle 'main trend'. Altering the degree of partial melting required to generate MORB has little effect on the relevant element ratios (Table 3). Having obtained a value of $f_{Sm/Nd}$ for the MORB source it is also possible to calculate the present-day isotopic composition of this source, assuming the present Sm/Nd ratio has been maintained since time t . For an age of 2,500 Myr the present day MORB source would have $\epsilon_{Nd} = 10.7$ and $\epsilon_{Sr} = -30$, values similar to those observed in recent basalts^{2,4,6}. This is at least consistent with the argument that Archaean crustal extraction may have been responsible for the widespread depletion of upper mantle sources.

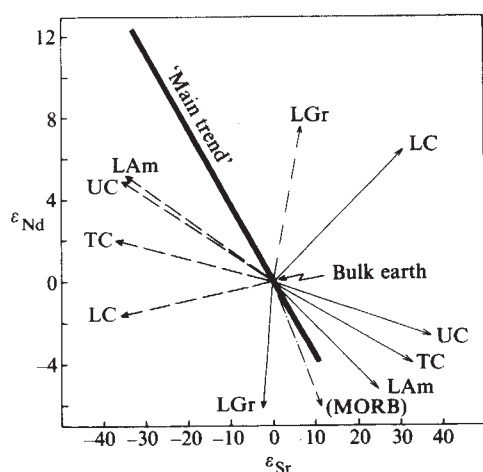


Fig. 3 Vectors showing the effect that extraction of different crustal components (Table 1, Fig. 2) would have on the Sr and Nd isotope composition of residual mantle which initially had bulk Earth Rb/Sr and Sm/Nd ratios and chondritic Sr/Nd ratios. —, Vectors for the crustal components themselves; ---, corresponding residual mantle vectors shown as dashed lines; -.-, the calculated crustal vector corresponding to the mantle source for typical mid-ocean ridge basalt.

The calculated complementary crustal component to this MORB mantle reservoir is detailed in Table 3 and defines a high (negative) slope vector on the $\epsilon_{Nd}-\epsilon_{Sr}$ diagram (Fig. 3), which lies between the crustal vectors corresponding to Lewisian granulite and amphibolite facies gneisses. The trend of this vector reinforces the conclusion that a lower crust component with a composition similar to that of Lewisian granulites is necessary for reasonable models of crustal growth. Note that the present Rb/Sr ratios of Lewisian granulites are not primary values, but have been lowered by Rb removal during the granulite-facies event, while the Lewisian amphibolite-facies gneisses may have suffered some corresponding Rb-addition (as is also implied by the Sr isotope data, ref. 12). Correcting for these effects would move both Lewisian crustal vectors closer to that corresponding to the MORB source.

Nevertheless, even accepting a higher bulk crustal Nd/Sm ratio than that of average andesite, it does seem that the bulk crust must have a fairly low mean Rb/Sr ratio, as concluded by DePaolo¹⁶. Using the R_t value for the bulk crust complement to the MORB source (Table 3) and a likely value of $f_{Sm/Nd}$ for the bulk crust (about -0.45 , Table 2) yields an $f_{Rb/Sr}$ value of ~ 1 for the bulk crust, corresponding to a mean Rb/Sr ratio of 0.07 . This value is low in relation to typical upper crustal values and must imply that a substantial proportion of the deeper continental crust has very low Rb/Sr ratios similar to those in Lewisian granulites.

Table 3 Calculated composition of MORB source and corresponding modelling parameters

	Average* MORB	MORB source†		
		$F = 0.10$	$F = 0.15$	$F = 0.20$
Rb	1.0	0.10	0.15	0.20
Sr	124	13.0	19.2	25.3
Nd	7.7	0.86	1.24	1.62
Sm	2.8	0.32	0.46	0.60
Sr/Nd	16.1	15.1	15.5	15.5
$f_{Rb/Sr}$	—	$-0.74\pm$	-0.73	-0.73
$f_{Sm/Nd}$	—	0.17	0.17	0.17
R_t	—	-0.24	-0.23	-0.23
ϵ_{Nd} (2,500 Myr)	—	10.7§	10.7	10.7
ϵ_{Sr} (2,500 Myr)	—	-30.5	-30.1	-30.1

* Average MORB from ref. 41.

† MORB mantle source compositions calculated assuming residual mineralogy: Ol 70, Opx 25, Cpx 5 and using partition coefficient data from ref. 42 and the equation:

$$C_{\text{source}} = C_{\text{MORB}}[D(1-F) + F]$$

where D = bulk distribution coefficient and F = fraction melted.

‡ R_t , $f_{Rb/Sr}$ and $f_{Sm/Nd}$ calculated as in Table 2.

§ ϵ_{Nd} and ϵ_{Sr} calculated from the relationship¹⁶:

$$\epsilon_{Nd}(0) \approx 24.7 f_{Sm/Nd} \cdot T \quad \text{and} \quad \epsilon_{Sr}(0) \approx 16.5 f_{Rb/Sr} \cdot T$$

Conclusions

The present modelling suggests that the Sr–Nd isotope correlation in the present upper mantle reservoir can be explained by extracting observed crustal compositions. However, crustal estimates based on Lewisian type crustal components seem to be more compatible with the isotopic constraints than crustal compositions based on the andesite model for crustal growth. Thus Archaean crust may, to a first approximation, complement the lithophile element depletion observed in modern MORB mantle sources. There are other interesting implications:

- (1) Quoted 'average' upper crustal compositions are unlikely to be representative of the whole crust.
- (2) The Precambrian crust, and particularly the lower crust, must have significantly higher Nd/Sm (and hence higher Ce_N/Yb_N) ratios than suggested by the andesite model for crustal growth.
- (3) The higher Ce_N/Yb_N ratios implied for (and the heavy REE depletion observed in) many Archaean crustal rocks suggest a role for eclogite in Archaean crustal generation processes. Because the residues of crustal extraction need to be recycled into the mantle to generate the Sr–Nd isotope correlation, this suggests that some form of subduction may have operated in the Archaean.
- (4) The requirement for high Ce_N/Yb_N ratios in the lower crust is not compatible with a lower crust dominated by the residues of partial melting or the cumulates from fractional crystallization.
- (5) Even allowing for higher Ce_N/Yb_N ratios in the bulk Archaean crust, it is apparent that a substantial proportion of the deeper continental crust must have a very low Rb/Sr ratio. This may have arisen through transfer of Rb and other radioactive heat-producing elements (U, Th, K) from the deeper crust to the upper crust during the Precambrian. A consequence is that the Precambrian lower crust may constitute a substantial reservoir which is unradiogenic with respect to Pb Isotope ratios as well as to Sr and Nd isotope ratios^{43,44}.

B.L.W. acknowledges receipt of an NERC studentship. We thank Drs C. J. Hawkesworth, M. J. Norry, R. K. O'Nions, A. D. Saunders and R. S. Thorpe for helpful comments.

Received 28 December 1979; accepted 15 May 1980.

1. Richard, P., Shimizu, N. & Allègre, C. J. *Earth planet. Sci. Lett.* **31**, 269 (1976).
2. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249 (1976).
3. O'Nions, R. K., Carter, S. R., Evensen, N. M. & Hamilton, P. J. *A. Rev. Earth planet. Sci.* **7**, 11 (1979).
4. De Paolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 743 (1976).

5. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **4**, 465 (1977).
6. O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13 (1977).
7. Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Earth planet. Sci. Lett.* **37**, 401 (1978).
8. Hawkesworth, C. J., Norry, M. J., Roddick, J. C. & Vollmer, E. *Nature* **280**, 28 (1979).
9. O'Nions, R. K., Evensen, N. M., Carter, S. R. & Hamilton, P. J. in *Results of Deep Sea Drilling in the Atlantic: Ocean Crust* (eds Talwani, M., Harrison, G. C. & Hayes, D. E.) Maurice Ewing Ser. **2**, 342 (American Geophysical Union, Washington, 1979).
10. Hamilton, P. J., Evensen, N. M., O'Nions, R. K., Smith, H. S. & Erlank, A. J. *Nature* **279**, 298 (1979).
11. Hamilton, P. J., O'Nions, R. K. & Evensen, N. M. *Earth planet. Sci. Lett.* **36**, 263 (1977).
12. Hamilton, P. J., Evensen, N. M., O'Nions, R. K. & Tarney, J. *Nature* **277**, 25 (1979).
13. McCulloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003 (1978).
14. Moorbath, S. *Nature* **254**, 395 (1975).
15. Moorbath, S. *Phil. Trans. R. Soc. A278*, 401 (1978).
16. DePaolo, D. J. *Earth planet. Sci. Lett.* **43**, 201 (1979).
17. DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 615 (1979).
18. Allègre, C. J., Ben Othman, D., Polve, M. & Richard, P. *Phys. Earth planet. Inter.* **19**, 293 (1979).
19. Nance, W. B. & Taylor, S. R. *Geochim. cosmochim. Acta* **41**, 225 (1977).
20. Sheraton, J. W., Skinner, A. C. & Tarney, J. in *The Early Precambrian of Scotland and Related Rocks of Greenland* 13 (eds Park, R. G. & Tarney, J.) (University of Keele, 1973).
21. Sheraton, J. W. *Earth planet. Sci. Lett.* **8**, 301 (1970).
22. Tarney, J., Skinner, A. C. & Sheraton, J. W. *24th Int. geol. Cong. Montreal* **1**, 162 (1972).
23. Tarney, J. in *The Early History of the Earth* 405 (ed. Windley, B. F.) (Wiley, London, 1976).
24. Tarney, J., Weaver, B. L. & Drury, S. A. in *Trondhjemites, dacites and related rocks* (ed. Barker, F.) 275 (Elsevier, Amsterdam, 1979).
25. Moorbath, S., Powell, J. L. & Taylor, P. N. *J. geol. Soc.* **131**, 213 (1975).
26. Moorbath, S., Welke, H. & Gale, N. H. *Earth planet. Sci. Lett.* **6**, 245 (1969).
27. Chapman, H. J. & Moorbath, S. *Nature* **268**, 41 (1977).
28. Pidgeon, R. T. & Aftalion, M. *Earth planet. Sci. Lett.* **17**, 269 (1972).
29. Pidgeon, R. T. & Bowes, D. R. *Geol. Mag.* **109**, 247 (1972).
30. Lyon, T. D. B. & Bowes, D. R. *Krystallikum* **13**, 53 (1977).
31. Hawkesworth, C. J., O'Nions, R. K., Pankhurst, R. J., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **36**, 253 (1977).
32. Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Science* **202**, 743 (1978).
33. Taylor, S. R. in *Island Arcs, Deep Sea Trenches and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) Maurice Ewing Ser. **1**, 325 (American Geophysical Union, Washington, 1977).
34. Hall, J. & Al-Haddad, F. M. *Scott. J. Geol.* **12**, 305 (1976).
35. Hall, J. & Simmons, G. *Geophys. J. R. astr. Soc.* **58**, 337 (1979).
36. O'Hara, M. J. *J. geol. Soc.* **134**, 185 (1977).
37. Weaver, B. L. & Tarney, J. (in preparation).
38. Tarney, J. & Windley, B. F. *J. geol. Soc.* **134**, 153 (1977).
39. Holland, J. G. & Lambert, R. St. J. *Precambrian Res.* **2**, 161 (1975).
40. O'Nions, R. K., Evensen, N. M., Hamilton, P. J. & Carter, S. R. *Phil. Trans. R. Soc. A258*, 547 (1978).
41. Sun, S.-S. & Nesbitt, R. W. *Earth planet. Sci. Lett.* **35**, 429 (1977).
42. Hanson, G. N. *J. geol. Soc.* **134**, 235 (1977).
43. Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *EOS* **60**, 406 (1979).
44. Tarney, J., Wood, D. A., Saunders, A. D., Cann, J. R. & Varet, J. *Phil. Trans. R. Soc.* (in the press).

Three-dimensional structure of *Escherichia coli* initiator tRNA_f^{Met}

Nancy H. Woo*, Bruce A. Roe† & Alexander Rich*‡

* Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

† Department of Chemistry, Kent State University, Kent, Ohio 44240

The crystal structure of Escherichia coli tRNA_f^{Met} an initiator transfer RNA, has been determined. While grossly similar to that of the chain-elongating yeast tRNA^{Phe}, there are three major differences. One involves the folding of the anticodon loop; in particular, the position of the constant uridine, U33. This difference was unexpected and may be of functional significance.

ALL organisms use a specialized transfer RNA molecule, tRNA_f^{Met}, to initiate protein synthesis. These tRNAs are distinguished from chain-elongating tRNAs in several ways, perhaps the most important of which is that the aminoacylated initiator tRNA goes directly into the ribosomal P site while all other aminoacylated tRNAs go initially to the ribosomal A site and only later in protein synthesis translocate to the P site^{1,2}. Despite significant differences in their nucleotide sequences, prokaryotic and eukaryotic initiator tRNAs are readily interchangeable in *in vitro* protein synthetic systems and thus form a single functional class. To understand the nature of the dis-

tinction between initiator and chain-elongating tRNAs, it is of value to compare the three-dimensional structure of an initiator tRNA with the known structure of the chain-elongating tRNA, yeast tRNA^{Phe} (refs 1-3). Here we report an X-ray diffraction analysis of the *Escherichia coli* tRNA_f^{Met}. We have crystallized this molecule in a form which diffracts to a resolution of 2.8 Å. The current analysis is based on data to 3.5 Å. The general shape of the molecule is similar to that of yeast tRNA^{Phe}, which has also been reported for yeast tRNA_f^{Met} (ref. 4). One of the differences from yeast tRNA^{Phe} is in the anticodon loop and that may be functionally significant in view of recent solution studies⁵.

‡ To whom correspondence should be addressed.

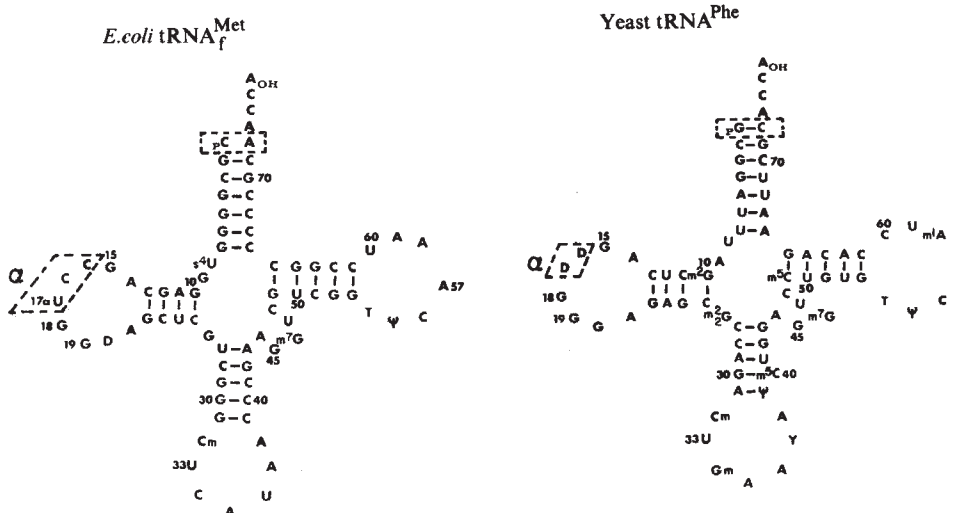


Fig. 1 The nucleotide sequence of the *E. coli* initiator tRNA, tRNA_f^{Met} and the chain-elongating yeast tRNA^{Phe}. The significant differences in the two sequences are indicated by the dashed boxes. The initiator tRNA contains unpaired bases at the end of its acceptor stem. The α region of the initiator tRNA contains one residue more than the corresponding region in yeast tRNA^{Phe}. The numbering system adopted here and in the text is that of Gauss *et al.*²³

Crystallization and structure solution

E. coli tRNA^{Met} was prepared as described by Roe *et al.*⁶, and Anandaraj and Roe⁷. The tRNA is crystallized in an orthorhombic lattice by vapour diffusion of 1–2% isopropanol into a solution containing 10 mM MgCl₂, 50 mM (NH₄)₂SO₄, 8 mM spermine and 2 mM BaCl₂ (ref. 8). Data to 3.5 Å (4,400 reflections) have been collected. Over 65% of the data in the shell 3.50–3.55 Å (217 reflections) diffract with an intensity significantly above background. Oligonucleotide binding, chemical modification, and NMR studies^{2,9} all suggest that no major conformational differences exist between yeast tRNA^{Phe} and *E. coli* tRNA^{Met}. In particular, no differences are found in the accessibility of nucleotides in *E. coli* tRNA^{Met}, corresponding to those involved in tertiary interactions in yeast tRNA^{Phe}, implying that the gross conformation of the two molecules is very similar. Consequently, the method of rotation and translation function searches was used to locate the initiator tRNA in its unit cell¹¹. The search model was the known crystal structure of the yeast tRNA^{Phe}.

The result of the searches was a unique position of the yeast tRNA^{Phe} search model in the *E. coli* tRNA^{Met} unit cell which formed a lattice without interpenetration of adjacent molecules. The initial model was formed by substituting the *E. coli* tRNA^{Met} base sequence into the yeast tRNA^{Phe} search model in such a way as to leave the conformation of the ribose-phosphate backbone intact. *E. coli* tRNA^{Met} contains one residue more in the variable α region of the D loop than yeast tRNA^{Phe} (see Fig. 1). The extra residue, C17, was not included in the initial model, nor was the 3' terminal adenosine 76. The agreement between a molecular model and the observed diffraction data is measured by the residual function R , defined as $R = \sum \|F_o\| - \|F_c\| / \sum \|F_o\|$ where F_o and F_c are the observed and calculated structure factors¹² which varies inversely with the agreement between model and data. The R value for the initial model was 0.44 at 6 Å, which compares favourably with residuals for protein models obtained using rotation-translation function search methods^{10,13}, as well as those obtained using multiple isomorphous replacement methods¹⁴. A complete description of the search procedure used is given in ref. 11.

Although the R for the initial model was encouraging and indicated that the hypothesis of gross similarity in conformation of *E. coli* tRNA^{Met} and yeast tRNA^{Phe} was correct, it also indicated that the model was far from perfect. This was further reflected by the R at 3.5 Å, which was 0.52. To determine if the disagreement was localized to some specific parts of the model, Fourier maps based on the initial model were calculated. At 3.5 Å resolution, both the difference Fourier, $F_o - F_c$, and the sum function Fourier, $2F_o - F_c$, showed that in most regions there was good agreement between the observed data and model but in some, the model was clearly wrong. These latter regions included the acceptor terminus, the D loop, and the anticodon loop. Fourier maps were also calculated from partial models in which those regions most at variance with the observed data were omitted; these included the entire anticodon loop, the entire D loop, the acceptor end. In both the $F_o - F_c$ and $2F_o - F_c$ maps calculated using the partial models, chains of continuous electron density were seen which could account for the omitted portion of the model. Thus the phasing from the remainder of the model was sufficiently accurate to fix the positions of the omitted portions.

The initial model was rebuilt using the University of North Carolina molecular graphics facility. Electron density maps ($F_o - F_c$ and $2F_o - F_c$) based on the complete model as well as those based on partial models were used to guide the fitting. The entire molecule was fitted to density and residue C17 inserted. The refit model was idealized using a restrained least-squares procedure¹⁵. The idealization involved both local geometry (bond angles and lengths) and hydrogen-bonding geometry of residues in the four stem regions. No attempt was made to preserve interactions between residues corresponding to those involved in tertiary bonding in yeast tRNA^{Phe} (refs 1–9). Limited

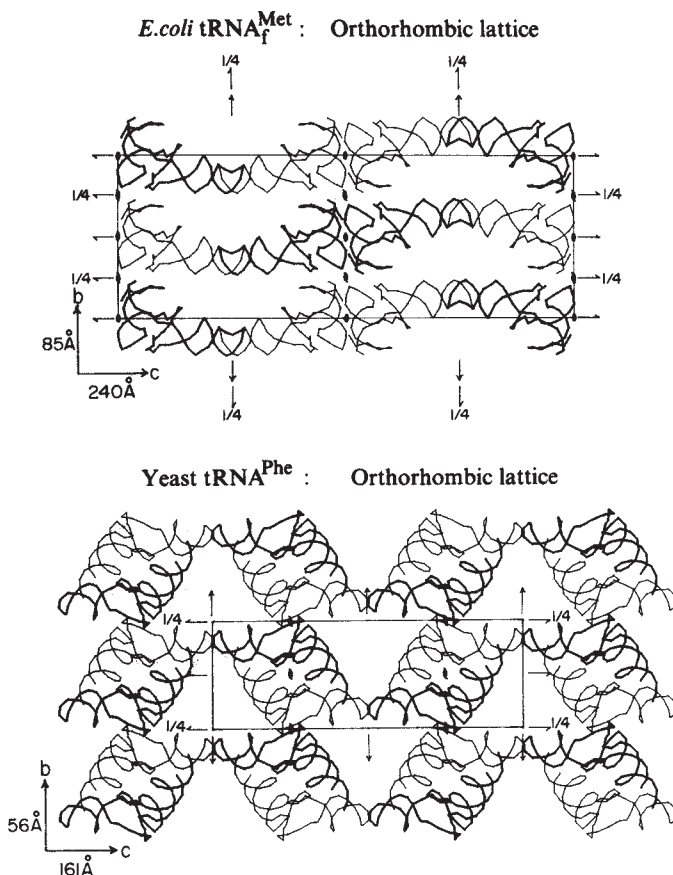


Fig. 2 Crystal packing diagrams of the *E. coli* tRNA^{Met} and yeast tRNA^{Phe}. The two packing diagrams are shown for comparative purposes. The diagrams show the projection of the molecular packing in the bc plane. Molecules which lie above are drawn in heavier lines than those below. Crystal symmetry elements are indicated. The empty regions are solvent channels which in both crystals extend through the crystal along the a axis direction. The intermolecular contacts in the two crystals can be seen to be totally distinct. The *E. coli* tRNA^{Met} crystals belong to the orthorhombic space group C222₁, with cell dimensions $a = 32$, $b = 85$, and $c = 240$ Å.

refinement of the idealized model has been refined together (for example, the anticodon loop and stem, the D and T loops). No refinement of the complete molecule at one time has been done. An isotropic temperature factor of 40 Å has been assigned to the molecule and has not been refined. The current resolution at 3.5 Å is 0.35. A fuller description will be given elsewhere (manuscript in preparation).

The lattice of *E. coli* tRNA^{Met} is shown in Fig. 2 with, for comparison, the orthorhombic lattice of yeast tRNA^{Phe}. Long solvent channels extend through the crystal along the a axis direction in both lattices. The lattice contacts in the initiator tRNA crystal include interactions between the T stems of twofold related molecules, between the acceptor terminus of one molecule with the T and D loops of a C-centred related one, and between the anticodon loops of twofold related molecules. These interactions are all distinct from those seen in either the orthorhombic or monoclinic lattices of the yeast tRNA^{Phe} (refs 16, 17). Hence, any similarity in the overall shape of the *E. coli* initiator tRNA and the yeast tRNA^{Phe} cannot be attributed to crystal lattice forces.

The form of the molecule

Figure 3 shows side views of both *E. coli* tRNA^{Met} and yeast tRNA^{Phe}. The two molecules are broadly similar in shape, the most striking difference being found in the acceptor ends. In the

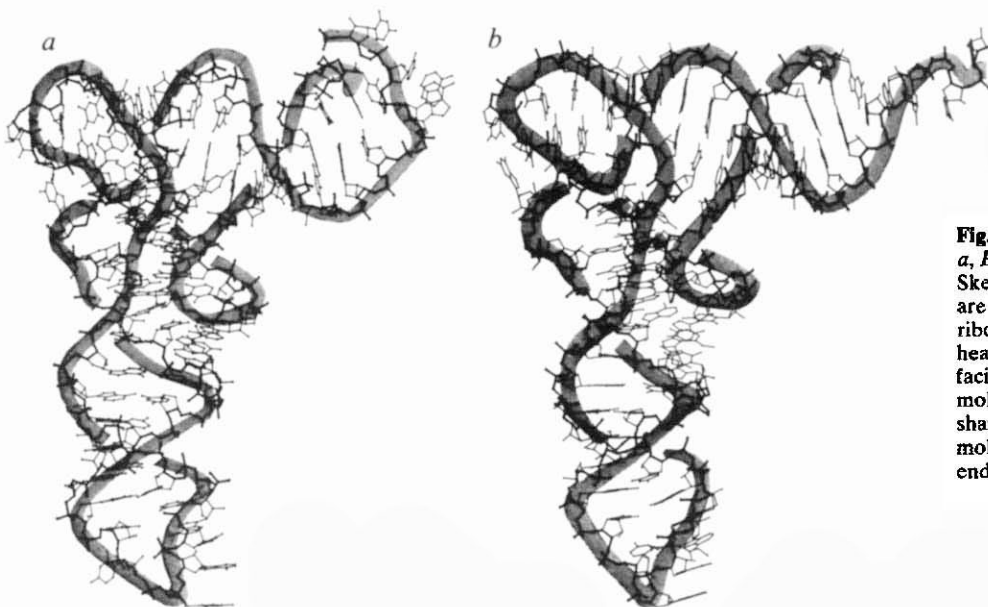


Fig. 3 A comparison of the structure of *a*, *E. coli* tRNA^{Met} and *b*, yeast tRNA^{Phe}. Skeletal diagrams of the two molecules are shown. The continuous phosphate-ribose backbone has been drawn in heavier lines and shading has been used to facilitate tracing of the chain. The two molecules are seen in the fully extended L-shape. The acceptor termini of both molecules are located at the upper right end of the L's and the anticodon loops at the bottom.

initiator tRNA, the last five nucleotides at the 3' terminus curl back toward the acceptor end, rather than continuing the helical organization of the acceptor stem, as in yeast tRNA^{Phe}. This difference in the acceptor ends of the two molecules may arise from the fact that the *E. coli* initiator tRNA, as most other prokaryotic initiator tRNAs, has an unpaired set of bases at the end of its acceptor stem (Fig. 1).

Several sections of the 3.5 Å electron density map based on the current model are presented in Fig. 4. The sum function Fourier map, $2F_0 - F_c$, is shown to allow assessment of both the quality of the fit and the noise level of the map. The phosphate-ribose backbone is found in strong electron density, as are virtually all of the bases. A few bases do not lie entirely in density, which indicates that their position will probably shift slightly with further refinement. The low density peaks found in solvent regions indicate the noise level of the maps. No solvent has been included in the current model. Figure 4*a* shows a portion of the map including the 5' side of the anticodon loop (see Fig. 4*e*). The position of uracil 33 differs in the initiator tRNA from that in yeast tRNA^{Phe} and will be discussed below. All of residue 33: phosphate, ribose, and base, lies in strong density. Figure 4*b* shows portions of the D stem and loop and the T loop. The D loop of the initiator tRNA was fitted using maps based on partial models to insert the extra residue in the α region. The base of residue 16 in the α region can be seen to lie in strong density, as do the bases of the D stem seen between phosphate 15 and ribose 23. The acceptor stem density is shown in Fig. 4*c*. The present model includes P76, the terminal phosphate, but not adenosine 76. The strong peak of density to the upper right of P76 is probably occupied by the ribose and the elongated extension may be the position of adenine 76. The strong density to the left of P76 is occupied by phosphates 73–75, which are part of the curled portion of the acceptor terminus (see Fig. 3). Figure 4*d* shows the anticodon arm at the point where the anticodon stem and D stem join to form one helix. At the left is a large region with a number of small peaks, which corresponds to the long solvent channel seen in Fig. 2. Figure 4*f* shows the D loop, T stem and acceptor stem. The generally good fit of the helix can be seen. Note the density corresponding to phosphate 73 which begins the curl seen in the acceptor terminus.

Figure 5 has stereo side and end views of *E. coli* tRNA^{Met}, and for comparison, yeast tRNA^{Phe}. In the side view of the initiator tRNA (Fig. 5*a*), one can see four of the five unpaired residues at the 3' end (A76 is not yet in the model). Residue 72 is base-paired to residue 1 in chain-elongating tRNAs and eukaryotic initiators^{1,2}; here it is rotated away from the base-pairing plane. The plane of adenine 72 is almost perpendicular to that of C1,

and is in a curved stacking array with bases 73, 74, and 75. The 3' terminus in this tRNA curves over the minor groove of the acceptor helix, bringing phosphate 76 close to phosphates 3 and 4 of the opposite strand. The significance of this conformation is not clear. Schulman *et al.*¹⁸ have shown that modification of C1 to U1, with the resulting formation of a normal Watson-Crick base pair with A72, does not interfere with initiator function, but does enhance the ability of the initiator tRNA to bind to *E. coli* elongation factor, Ef-Tu. The curved stacking array of bases 72 through 75 appears to be stable in the crystal. It would be of interest to determine its stability in solution.

Although eukaryotic initiators differ from other tRNAs in the T loop, prokaryotic initiators differ in this loop only in the presence of A57 rather than the more common G57^{1,2}. The major difference seen here in the initiator tRNA structure compared to yeast tRNA^{Phe} is the position of residue 57. A57 in the initiator tRNA is intercalated between G18 and G19 of the D loop, as is G57 in yeast tRNA^{Phe}, but is not drawn in as far as the backbone of the D loop (compare Fig. 5*a* and *c*). The position of G57 in yeast tRNA^{Phe} may be stabilized by the hydrogen bonds formed between N2 of G57 and riboses 18 and 19 (refs 19, 20). Such stabilizing forces are not possible when adenine is substituted for guanine, which may account for the differences. The current model of the initiator tRNA has not been constrained to include any tertiary interactions. The only constraints imposed have been on residues involved in the hydrogen-bonded stem regions. Most of the base positions in the current initiator tRNA model have been shifted from their original positions in the initial model by the refitting process. Tertiary interactions cannot be identified with certainty at the resolution of this report. However, they can be tentatively identified by the relative geometry and proximity of the bases. For example, O6 of G9 appears to be within hydrogen-bonding distance of the amino group of C12, an interaction which may be similar to that in yeast tRNA^{Phe} between A9 and A23. Most of the tertiary interactions seen in yeast tRNA^{Phe} appear also to exist in *E. coli* initiator tRNA. The precise identification of these interactions as well as the identification of the interactions of magnesium ions and spermine must, however, await the higher resolution analysis.

The major differences between the two structures are found outside the core region of the molecule. There are two significant sequence differences between *E. coli* tRNA^{Met} and yeast tRNA^{Phe} (Fig. 1). The first is found at the end of the acceptor stem and may give rise to the difference in conformation of the 3' OH end of the two molecules described earlier. The second is the size of the α region of the D loop. This region of the initiator structure was fit using Fourier maps based

on a model in which the entire D loop was omitted. The α region residues are shown shaded in Fig. 5a, c. In the initiator tRNAs these residues are associated with strong density (see Fig. 4c) and appear to be more rigidly fixed than the two α residues in yeast tRNA^{Phe}. D16 and D17 in yeast tRNA^{Phe} point away from each other, while the three α bases in the initiator tRNA are close together, with C17 and U17a stacked upon each other (Fig. 5b). Although the α region in this initiator tRNA is one residue larger than that of yeast tRNA^{Phe}, it does not project further from the body of the molecule than the α region of yeast tRNA^{Phe} (Fig. 5b, d). Instead, the D loop accommodates the extra residue by folding toward the body of the molecule, giving rise to a more tightly organized loop than in yeast tRNA^{Phe}. This may explain why tRNAs have a maximum of three nucleotides in the α region^{1,2,8}; there is simply not enough room for an internally directed segment to accommodate a fourth nucleotide without encroaching on the body of the molecule.

The anticodon loops of *E. coli* tRNA^{Met} and yeast tRNA^{Phe} are superficially similar in that the anticodon bases have roughly similar stacking (Figs 3 and 5a, b). However, an important difference lies in the orientation of the conserved residue U33 which is 5' to the anticodon. In yeast tRNA^{Phe}, U33 folds into the anticodon loop and forms a hydrogen bond between N3 of U33

and phosphate 36 (refs 19, 20). U33 in *E. coli* tRNA^{Met} has an almost opposite orientation; it is thrust away from the anticodon loop into the solvent region. An intermolecular lattice interaction involving uracil 33 may help stabilize its orientation; however, its identification is still tentative. The effect of this difference in U33 orientation on the anticodon loop folding in the two molecules can be seen in Fig. 5b, d. The loop in yeast tRNA^{Phe} is symmetrical while the loop of *E. coli* tRNA^{Met} is skewed, so that the anticodon bases are in different positions relative to the anticodon stem in the two molecules. The orientation of residue 33 in the initiator tRNA involves a position of the ribose so that the 2' hydroxyl of ribose 33 could perhaps form a hydrogen bond with phosphate 36. Phosphate 36 is also in a slightly different position than in yeast tRNA^{Phe}, being closer to the backbone of the 5' side of the anticodon loop. Residues 34 and 35 are forced off to one side, producing the skewed shape of the loop. The shape of the anticodon loop may thus be determined by the group hydrogen bonding to phosphate 36: one conformation of the loop is found if the hydrogen-bonding group is ribose 33 and another if it is uracil 33. The two conformations show a marked difference in the position of phosphate 35 which in the initiator tRNA extends away from the anticodon loop in contrast to its more proximal position in the yeast tRNA^{Phe}.

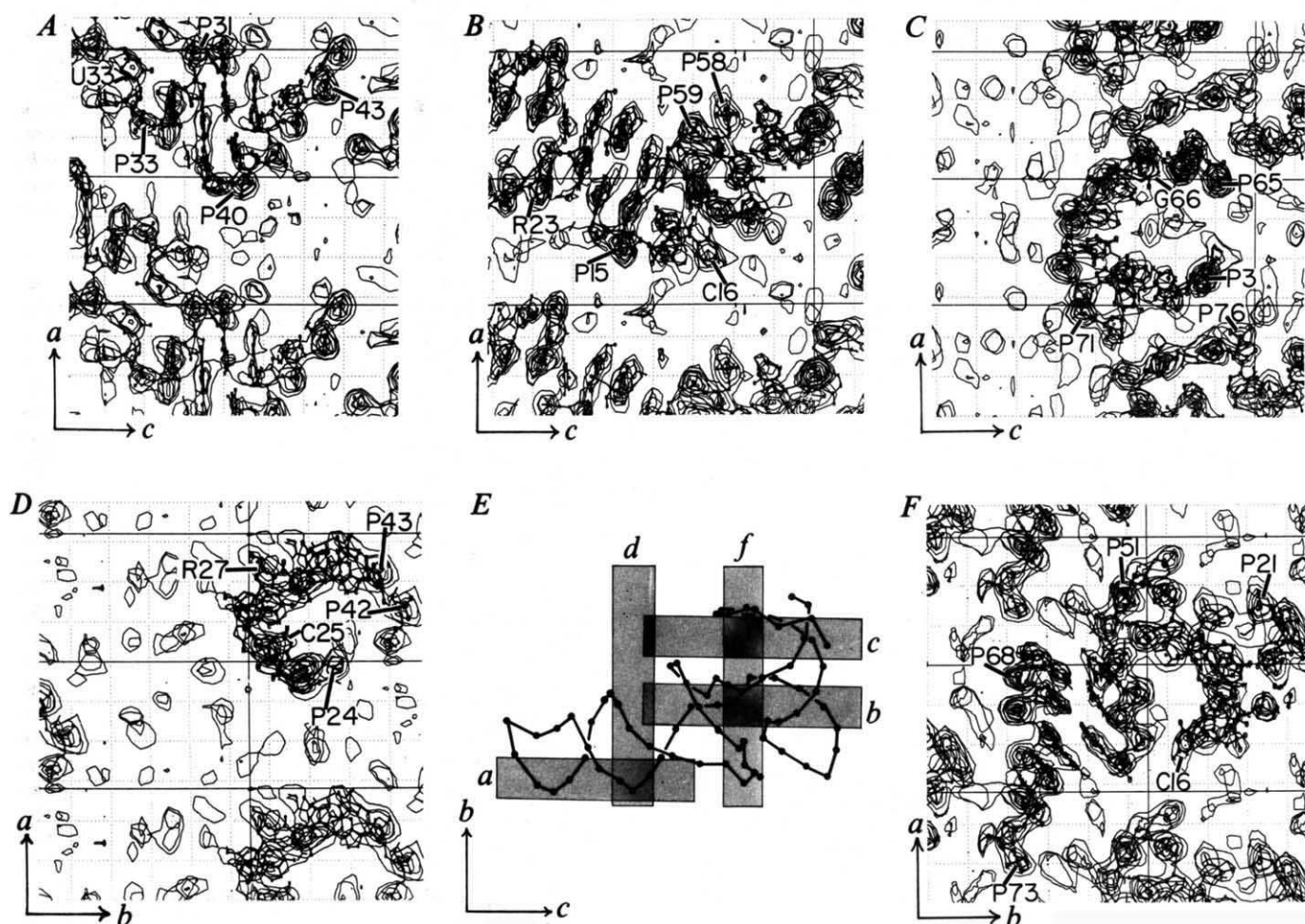


Fig. 4 Five different sections of the sum function Fourier map, $2F_0 - F_c$, are shown. Each section is the superposition of seven sections with a total thickness of 8 Å. In e, a schematic diagram is given which includes a plot of the phosphorous atoms in the backbone. The molecule is shown in the orientation of Fig. 2. The shaded bars indicate the regions of the molecule seen in the correspondingly labelled density section. The bars are all 8 Å wide and extend over 50 Å in the direction perpendicular to the page. The scale of this schematic is not the same as that of the density sections. In a-d and f, one-and-a-half unit cells are shown in the a direction. The horizontal lines are drawn at $x = 0, \frac{1}{2},$ and 1. The distance between each pair of lines is therefore 16 Å. In b and c, the vertical line is drawn at $z = 1/2$, and in d and f, at $y = 1/2$. In all sections the model has been drawn in addition to the contours of electron density. The principal molecule is drawn in heavy lines, symmetry-related molecules in light lines. The contouring begins at 0.275 e Å^{-3} with intervals of 0.300 e Å^{-3} . In indicating various parts of the molecule, we have used P for phosphate, R for ribose, and A, G, C, U for bases.

Fig. 2

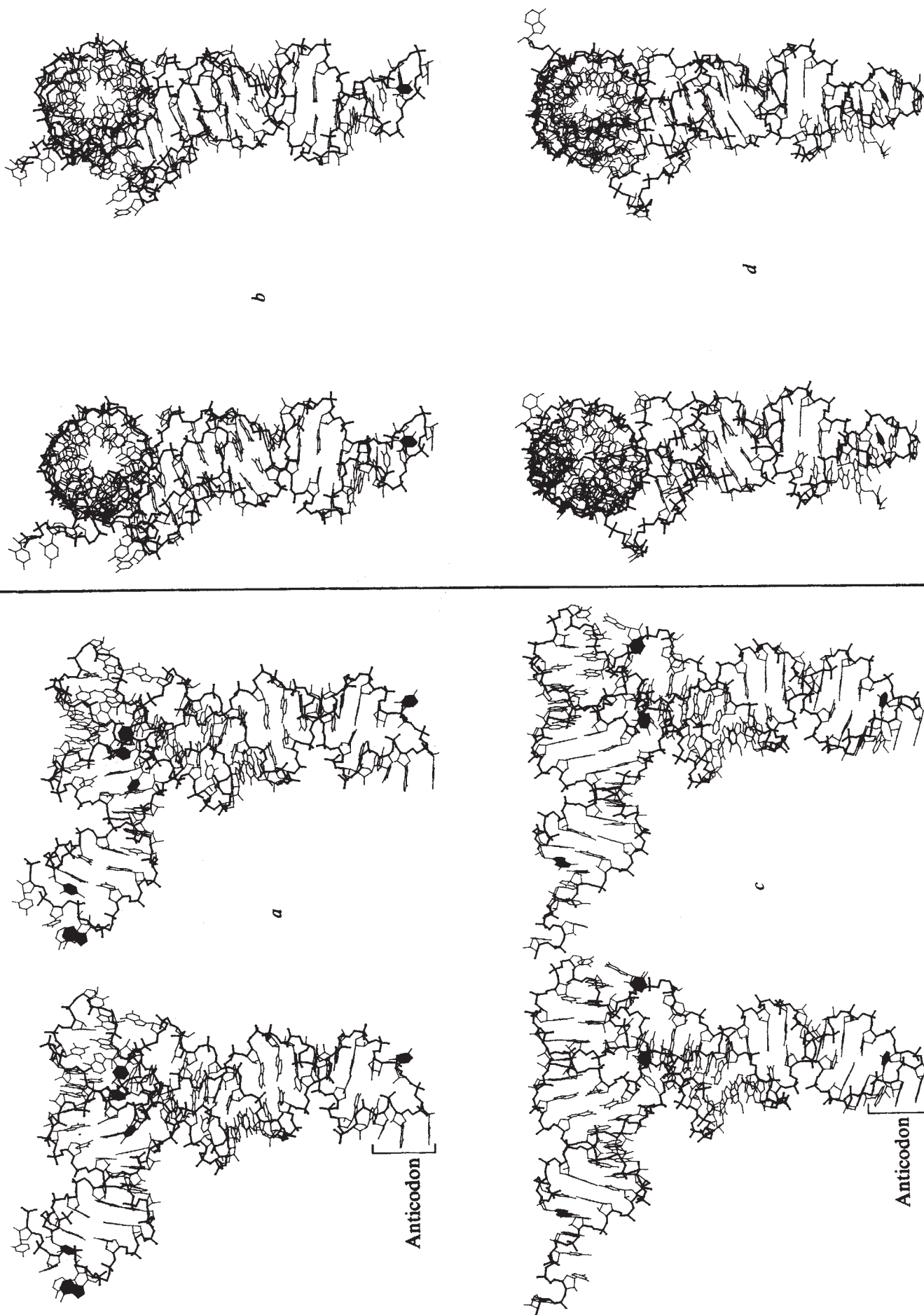


Fig. 5 Stereo diagrams of *E. coli* tRNA^{Met} (a, b) and yeast tRNA^{Phe} (c, d). a and c show side views of the two molecules and b and d views looking down the acceptor helix. In a, residues C1, A72, C16, C17, U17a, and U33 are shaded. Only residue U33 is shaded in b and d. In c, residues G1, C72, D16, D17, and U33 are shaded. The backbone of each molecule is emphasized with heavy lines.

Two different kinds of molecules

The structure presented here is based on data to 3.5 Å resolution and has been partially refined to a residual of 0.35. We have been able to interpret the structure at this resolution in more detail than might be expected because we have interpreted it comparatively with the refined 2.5 Å structure of the yeast tRNA^{Phe}. There are three regions in which the structure of the *E. coli* tRNA^{Met} differs substantially from that of yeast tRNA^{Phe}. The first is the 3' end of the molecule. The conformation observed in the initiator tRNA demonstrates the flexibility of this end of tRNA molecules which is consistent with its biological role: the transfer of the nascent polypeptide chain from one tRNA to the next^{1,2,9}. The second difference is found in the α region of the D loop. The initiator tRNA accommodates its larger α region not by extending it further into solvent, but by drawing it closer to the body of the molecule, giving a more compact structure than seen in yeast tRNA^{Phe}. The third difference concerns the conformation of the anticodon arm. In their 4.5 Å analysis of yeast tRNA^{Met}, Shevitz *et al.* found an overall similarity in the shape of the yeast initiator tRNA to that of yeast tRNA^{Phe}, with a difference existing in the anticodon arm⁴. It will be of interest to compare the structures of the *E. coli* and yeast initiator tRNAs once the details of the latter are available. A recent biochemical study has implicated the anticodon loop conformation in distinguishing initiator tRNAs from chain elongation tRNAs by their endonuclease cleavage patterns⁵. In the present study, the major difference in the anticodon region concerns itself with the orientation of uridine 33. In *E. coli* tRNA^{Met}, the base projects away from the loop while in yeast tRNA^{Phe}, it points into the loop^{19,20}. The effect of the different orientation is a position of ribose 33 which appears to allow the formation of a hydrogen bond between O2' of the ribose with phosphate 36. It is important to emphasize, however, that the identity of this hydrogen bond, as of all the tertiary hydrogen bonds discussed above, is tentative until the higher resolution analysis is complete. Phosphate 36 is drawn toward the centre of the loop while phosphate 35 is forced away from the centre, giving rise to the skewed appearance of the loop (Fig. 5b) in contrast to that of yeast tRNA^{Phe} (Fig. 5d). The two loop conformations lead to a slightly different stacking arrangement of the anticodon bases. Although the position of uracil 33 in *E. coli* tRNA^{Met} may be stabilized by a lattice interaction, it is not likely that the loop conformation is entirely determined by this interaction.

The difference in the anticodon loop conformation was not anticipated from either the nucleotide sequence or the results of

earlier solution studies. The implication from the recent nuclease digestion study⁵, however, is that the conformation detected there for the *E. coli* initiator tRNA is the same as that found in both lower and higher eukaryotic initiators. The generality of an initiator tRNA anticodon loop conformation may be the determining factor which permits initiators to enter the bind at the P site of the ribosome while chain-elongating tRNAs bind initially at the A site. The external position of uracil 33 and its concomitant failure to participate in hydrogen bonding within the anticodon loop offers an explanation for the cytosine in position 33 in higher eukaryotic initiators. In cytosine, N3 is not protonated so that a hydrogen bond between the base of residue 33 and phosphate 36 of the type seen in yeast tRNA cannot exist. However, even with cytosine in position 33, the hydrogen bond between ribose 33 and phosphate 36 suggested in the *E. coli* initiator tRNA could exist. One should recall here that higher eukaryotic initiators can be used *in vitro* *E. coli* protein synthetic systems and vice versa²¹.

The uridine swivel hypothesis

tRNAs move from the A site to the P site in the ribosome. The conformational difference in the anticodon loop which we observe between the chain-elongating yeast tRNA^{Phe} and the initiator *E. coli* tRNA^{Met} may be a conformational shift which occurs when tRNA moves from one site to the other. A chain-elongating tRNA moving from the A site to the P site may interact with an element of the ribosomal machinery which flips the anticodon loop from one conformation to the other with a swivel motion of U33. If so, does the externally oriented uracil 33 play a structural role when the tRNA occupies the P site? Perhaps a ternary complex forms involving messenger RNA and tRNAs in the A and P sites, such that uracil 33 of tRNA hydrogen bonds to the anticodon arm of the A-site tRNA. This hydrogen bond could involve a 2' hydroxyl of the A-site tRNA and O2 of uracil 33. In such an hypothesis, cytosine in position 33 would not modify the stability of the system. Furthermore, the hypothesis allows a plausible positioning of the tRNAs in the A and P sites in which the anticodon bases hydrogen bond to adjacent nucleotide triplets on the messenger RNA, and the CCA ends of the two molecules are close enough to effect a transfer of the nascent polypeptide chain. In this arrangement, the two codons of the message must be unstacked as suggested a few years ago²². A further description of this proposed ternary complex will be published elsewhere. The observed difference in the anticodon loop conformation of the *E. coli* initiator tRNA and the yeast tRNA^{Phe} may thus provide clues as to the manner in which tRNAs associate with each other in the ribosome.

This research was supported by grants from the NIH, NSF, NASA, and the American Cancer Society. We thank Dr F. P. Brooks and his colleagues at the Department of Computer Science, University of North Carolina at Chapel Hill, for their assistance with the GRIP-75 molecular graphics facility. We also thank Dr Gary Quigley for use of the program to plot Figs 2–5 and the coordinates in Figs 2 and 5, and both him and Andrew H.-J. Wang for their encouragement, advice, and assistance.

Received 8 October 1979; accepted 3 April 1980.

- Sigler, P. B. *A. Rev. Biophys. Bioengng* **4**, 477–527 (1975).
- Rich, A. & RajBhandary, U. L. *A. Rev. Biochem.* **45**, 805–860 (1976).
- Quigley, G. J., Teeter, M. M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 64–68 (1978).
- Shevitz, R. W. *et al. Nature* **278**, 188–190 (1979).
- Wrede, W., Woo, N. H. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3289–3293 (1979).
- Roe, B. A., Marcu, K. & Dudoock, B. *Biochim. biophys. Acta* **319**, 25–36 (1973).
- Anandaram, M. P. J. S. & Roe, B. A. *Biochemistry* **14**, 5068–5073 (1975).
- McPherson, A. Jr in *Methods of Biochemical Analysis* (ed. Glick, D.) 249–345 (Wiley, New York, 1976).
- Goddard, J. P. *Prog. Biophys. molec. Biol.* **32**, 233–308 (1977).
- Rossmann, M. G. (ed.) *The Molecular Replacement Method* (Gordon and Breach, New York, 1972).
- Woo, N. H. thesis, MIT (1978).

- Stout, G. H. & Jensen, L. H. in *X-ray Structure Determination*, 245 (Macmillan, New York, 1968).
- Ward, K. B., Wishner, B. C., Lattman, E. E. & Love, W. E. *J. molec. Biol.* **98**, 161–179 (1975).
- Blundell, T. L. & Johnson, L. N. in *Protein Crystallography*, 428–429 (Academic, London, 1976).
- Hedrickson, W. A. & Konnert, J. in *Biomolecular Structure, Conformation, Function, and Evolution* Vol. I (ed. Striniwason, R.) 43–57 (Pergamon, Oxford, 1979).
- Kim, S. H. *et al. Science* **185**, 435–440 (1974).
- Robertus, J. D. *et al. Nature* **250**, 546–550 (1974).
- Schulman, L. H., Pelka, H. & Sundari, R. M. *J. biol. Chem.* **249**, 7102–7110 (1974).
- Quigley, G. J. & Rich, A. *Science* **194**, 796–806 (1976).
- Jack, A., Ladner, J. E. & Klug, A. *J. mol. Biol.* **108**, 619–649 (1976).
- Elson, N. A., Adams, S. L., Merrick, W. C., Safer, B. & Anderson, W. F. *J. biol. Chem.* **250**, 3074–3079 (1975).
- Rich, A. in *Ribosomes*, 871–884 (Cold Spring Harbor Laboratory, New York, 1974).
- Gauss, D. G., Gruter, F. & Sprinzl, M. *Nucleic Acids Res.* **6**, r1–r19 (1979).

Insertion of the eukaryotic transposable element Ty1 creates a 5-base pair duplication

Philip J. Farabaugh & Gerald R. Fink

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

The his4-912 mutation results from insertion of a transposable element into the 5'-non-coding region of the his4 gene of yeast. A duplication of 5 base pairs of wild-type his4 DNA flanks the inserted element. The major class of His⁺ revertants result from excision of most of the transposable element, leaving an inserted segment of 334 base pairs flanked by the 5-base pair repeats.

TRANSPOSABLE genetic elements in bacteria are capable of transposition to new sites and promoting chromosomal rearrangements including deletions, inversions and fusions^{1,2}. The existence of eukaryotic transposable elements has been suggested on genetic and biochemical grounds. In maize, the genetic 'controlling elements' are capable of transposition to new sites and of catalysing gross rearrangements of the genome³. Reiterated DNA elements which have the same structure as a bacterial transposon have been isolated from *Drosophila*⁴ and the yeast *Saccharomyces cerevisiae*⁵. The location of these elements varies between strains. The genetic properties of maize controlling elements and the biochemical properties of the *Drosophila* and yeast elements suggest that transposable elements similar to those found in prokaryotes exist in eukaryotes. In no case has an unstable mutation been shown biochemically to be the result of the insertion of a transposable element.

The *his4-912* mutation is due to the insertion of a transposable element (Ty1) into the *his4* locus of yeast⁶. The mutation is unstable and has given rise to translocations, deletions, a transposition and an inversion⁷. DNA sequence analysis of the wild-type *his4* gene and the *his4-912* mutant indicates that transposition of Ty1 creates a 5-base pair duplication of DNA previously at the site. Excision of the element by recombination between the terminal 334-base pair repeats results in reversion to a His⁺ phenotype. In *his4-912*, Ty1 inserted into the 5'-non-

coding region of the *his4* gene. The phenotype of the His⁺ revertants of *his4-912* suggests that Ty1 does not disrupt the *his4* promoter but that it interferes with *his4* expression by transcriptional competition.

Analysis of His4⁺ and his 4-912

The *his4* locus was cloned as a 20×10³-base pair BamHI fragment on pBR313 (YIP300)⁸. Existing deletions of portions of *his4* localize the beginning of the gene to 500 base pairs from the end of an approximately 1,500-base pair SalI fragment⁹. The Southern blot of labelled *his4* DNA to DNA from a 912 yeast strain and the restriction map of the 912 clone demonstrated that the 912-insert was a Ty1 element transposed to the same SalI fragment⁶.

We present here the DNA sequence from this SalI fragment in His4⁺ and compare it with the DNA sequence from the SalI fragment in *his4-912*. We analysed fragments of the His4⁺ segment which end at SalI, HinfI, TaqI, MspI, XhoI and PvuII sites to determine the complete nucleotide sequence of the wild-type 1,577-base pair fragment (Fig. 1). The *his4* coding frame extends through the last 474 base pairs of the top strand of the fragment. Many nonsense codons are dispersed throughout the region in the other two coding frames. Two early *his4*⁻ mutations differ from the wild-type sequence within this segment (T. Donahue, personal communication). The protein specified by the *his4* coding frame begins Met-Val-Leu-Pro-Ile; the amino terminus of the *his4* protein is the tetrapeptide Val-Leu-X-Ile (X was unidentified, ref. 10 and R. Bigelis, personal communication). These data show that the coding region for the amino-terminal portion of the *his4* protein begins at position +1 shown in Fig. 2.

The plasmid containing the *his4-912* mutation (YIP312) includes 19×10³ base pairs of yeast DNA (the restriction map, as previously determined⁶, is shown in Fig. 3). To analyse more easily the DNA sequence in the region of the two junctions between *his4* and the insert, we cloned the two SalI fragments which span these junctions in pBR322 (YIP312-1 and YIP312-2; Fig. 3). The junction closest to *his4* was sequenced from fragments produced by cleavage of YIP312-1 with SalI, HinfI, Sau3A, TaqI and BglII. The junction distal to *his4* was sequenced using the restriction sites for the enzymes HinfI, MspI and Sau3A, which lie within a 900-base pair PvuII fragment of YIP312-2 (Fig. 4b). The sequence of the *his4*-proximal junction extends 715 base pairs into the insertion element and the *his4*-distal sequence extends 369 base pairs in from the other junction (Fig. 2).

Ty1 is a transposable element

The Ty1 element is a 5,600-base pair region flanked by direct repeats of approximately 250 base pairs ('δ regions')⁵. The element is present in about 35 copies dispersed throughout the

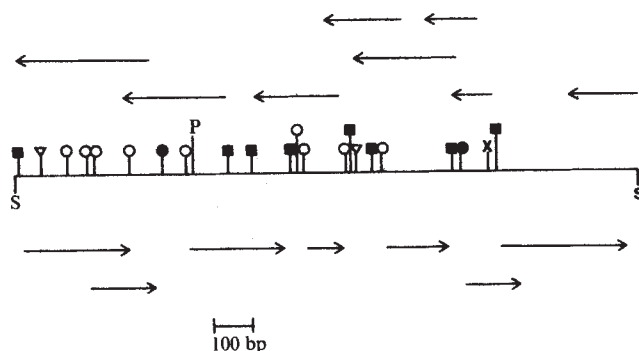


Fig. 1 The sequencing strategy for the 1,577-base pair (bp) His4⁺ SalI fragment. Those restriction sites used to determine the sequence are indicated: SalI(S), HinfI(H), TaqI(T), MspI(M), XhoI(X), Sau3A(A) and PvuII(P). Each arrow above and below the map indicates the extent of the 3' to 5' sequence determined from a single experiment; the head of the arrow indicates the position of the nucleotide identified farthest from the end label. The restriction fragments for sequence analysis were generated by cleavage of either the SalI plasmid (YIP302) or the SalI fragment and labelled at their 3' ends with either [α -³²P]dATP or dCTP using reverse transcriptase²⁹. The labelled fragments were isolated and subjected to the Maxam-Gilbert chemical sequencing procedure and the reaction products separated by electrophoresis as described³⁰ except that 80-cm gels were used to enable us to read sequences very far from the label ($\pm$ 350 base pairs). The products were visualized by autoradiography and the sequence read from the film.

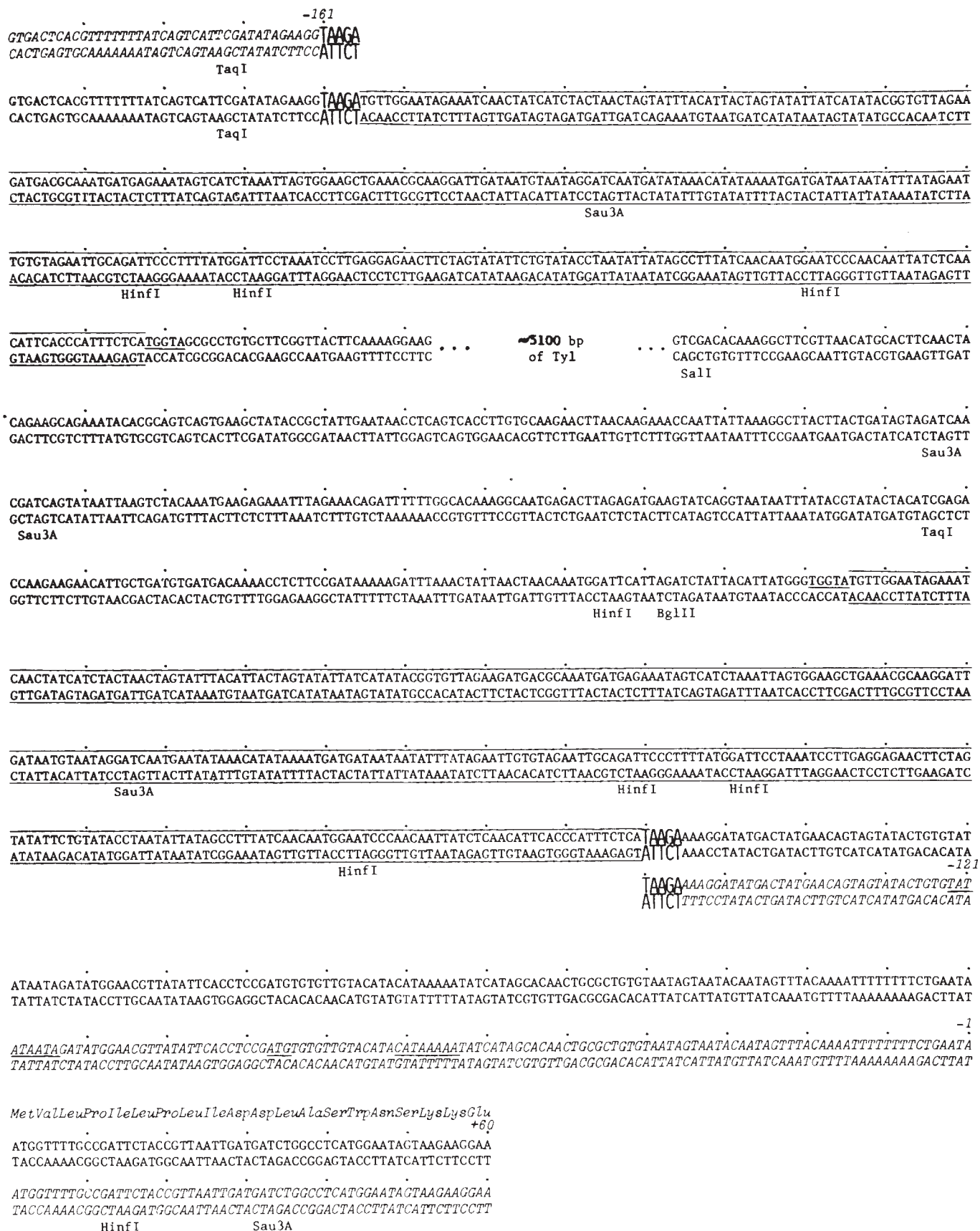


Fig. 2 Comparison of the *His4*⁺ and *his4*-912 sequences indicates the location of *Ty1*. The *his4*-912 sequence (roman letters) is homologous with the *His4*⁺ sequence (italic letters) up to position -161 from the right (the direction to *his4*) and -157 from the left. The sequence of 5 base pairs from positions -157 to -161 (large letters) is duplicated at either end of the element. The 334-base pair region into the element from either junction is a direct repeat of the δ element (indicated by bars above and below the *Ty1* sequence). The 5,600-base pair internal region of *Ty1* lies between the two δ elements. The sequence of 36 base pairs in from the right hand and 382 base pairs in from the other have been determined. At the junctions between the δ elements and the 5,600-base pair region there is a second 5-base pair duplication 5'-TGGTA-3' (underlined). The non-coding region of *His4*⁺ (italics) includes two possible Hogness boxes (underlined) that could be associated with mRNAs beginning on the *his4*-side of the first ATG in the *his4* non-coding region (underlined).

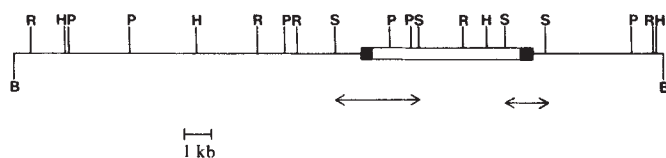


Fig. 3 The restriction map of the *his4-912* plasmid (YIP312) as described by Roeder and G.R.F.⁵. The sites for the enzymes *Bam*HI(B), *Eco*RI(E), *Hind*III(H), *Pst*I(P) and *Sal*I(S) are indicated. The boxed region represents the *Ty1* insert, and the δ repeats are indicated by the solid regions. The portions of this plasmid cloned in YIP312-1 and YIP312-2 are indicated by double-headed arrows below the map.

yeast genome and the δ regions are present in over 100 copies both as part of *Ty1* elements and as independent units. The *Ty1* element was inferred to transpose because it appeared in different locations in several strains of yeast. However, in no case was transposition directly observed.

The sequence of *his4-912* demonstrates that *Ty1* transposed into *his4*, as only the element itself has moved into the new site. The element which has inserted into *his4* is approximately 5,600 base pairs flanked by a perfect 334-base pair direct repeat. The *his4* wild-type sequence begins precisely at the ends of the terminal repeats, so no DNA other than *Ty1* has been inserted. Because insertion includes only *Ty1*, it could not have been caused by homologous recombination involving sequences flanking its original site. The DNA sequence of the *his4-912* mutation is consistent only with transposition of an intact *Ty1* into the 5'-non-coding region of *his4*.

The *Ty1* element in *his4-912* is flanked by a duplication of 5'-TAAGA-3' (Fig. 5). This duplication was created during transposition, because the sequence appears only once in *His4*⁺ DNA (nucleotides -161 to -157). The creation of flanking duplications is a characteristic of all prokaryotic transposable elements; both 5-base pair (IS2, Tn3, Mu and γ - δ) and 9-base pair duplications (IS1, Tn5, Tn9, Tn10) have been observed (summarized in ref. 11). Our finding that *Ty1* insertion creates a duplicated sequence in the target suggests that the mechanism of transposition is similar between a eukaryote and the prokaryotes.

A feature of the *Ty1* sequence which we cannot explain is the existence of a 5-base pair duplication of 5'-TGGTA-3' at the internal δ junctions of *Ty1* (Fig. 2). The existence of a duplication of the sort created by *Ty1* transposition suggests that these junctions have also been created by transposition. There is no simple transposition process known or postulated that could account for this structure. It is also possible that this is either a random occurrence of a 5-base pair duplication or that the duplication is required for *Ty1* function.

The insertion of *Ty1* creates a 24-base pair inverted repeat

Although in *Escherichia coli* the insertion of transposable elements is not a site-specific event, the sites of insertion cluster in specific regions ('regional specificity') and many insertion mutations are found in these regions at a few specific sites ('hotspots'). Many transposons demonstrate regional specificity [Tn3 (ref. 12), Tn7 (ref. 13), Tn9 (ref. 14), Tn10 (refs 15, 16), Tn501 and Tn802 (ref. 17)]. DNA sequencing of Tn10 (ref. 18) and Tn3 (ref. 19) insertions shows directly that these elements have insertion hotspots. The hotspots for insertion of Tn3 occur near a sequence in the target DNA which is homologous to the termini of the element. The insertion of Tn3 at the hotspots creates an inverted repeat between the homologous sequences in the element and the site. Thus, some sequence homology is responsible for the specificity of insertion, although homologous recombination is not involved.

A sequence created by the insertion of *Ty1* into *his4* suggests that transposition is directed by limited sequence homology. The sequences 12 base pairs in either direction from one mutation junction form an inverted repeat with its centre of symmetry precisely at the junction (Fig. 5). The sequence, 5'-ACC-

CATTCTCA/TAAGAAAAGGAT-3' (the slash is the centre of symmetry and the junction), includes nine matches and two purine-purine matches. The sequence at the end of δ could be required for recognition of the preferred insertion site. The recognition does not involve homologous recombination, as that process would create a flanking direct repeat.

Excision of *Ty1*

The major class of *His*⁺ revertants is caused by a sequence alteration at the 912 site. These revertants are *His*⁺ at 37 °C and *His*⁻ at 23 °C (ref. 7). The *his4* region from these revertants has suffered a large deletion of *Ty1* DNA, but still retains approximately 300 base pairs of the element⁶. The fact that the δ element is the same size as the remaining insert suggested that reversion resulted from recombination between the direct terminal repeats of δ , leaving a single δ . This process is called excision. The Tn9 element, which is flanked by a direct repeat of IS1, also undergoes an excision of this sort²⁰.

We cloned the *his4* region of one cold-sensitive revertant, *his4-912R1*, using a modification of a previously described procedure⁶. Sequence analysis of the *his4* region from *his4-912R1* shows that the mutation results from the presence of a single δ element which is identical to those at the end of *Ty1* in *his4-912*. The δ element is flanked by the identical 5-base pair duplication flanking *his4-912*. We have sequenced 250 base pairs towards and into the *his4* gene and found no additional differences from *His4*⁺. The sequence is entirely consistent with the recombination model for excision. Excision of *Ty1* by this specialized recombination could be responsible for the distribution of δ sequences in the yeast genome. As many as 50 δ sequences independent of *Ty1* are dispersed throughout the genome. These δ elements could mark the sites of previous *Ty1* excisions.

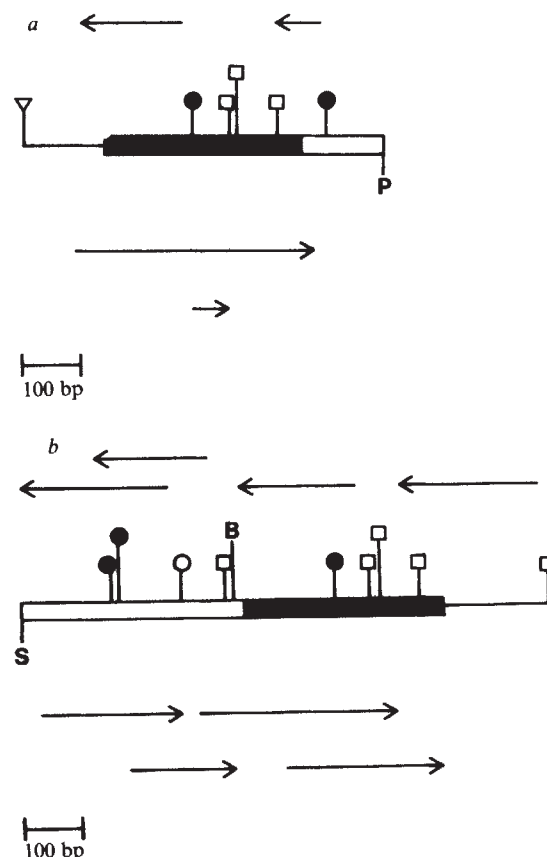


Fig. 4 The sequencing strategy for the two junctions of *Ty1* in *his4-912* is shown as described in Fig. 1 legend; the site for the enzyme *Bgl*II is indicated (B). The *Ty1* element is represented as in Fig. 3. *a*, The strategy for the *his4*-distal (left) region from plasmid YIP312-1. *b*, The strategy for the *his4*-proximal (right) region (YIP312-2).

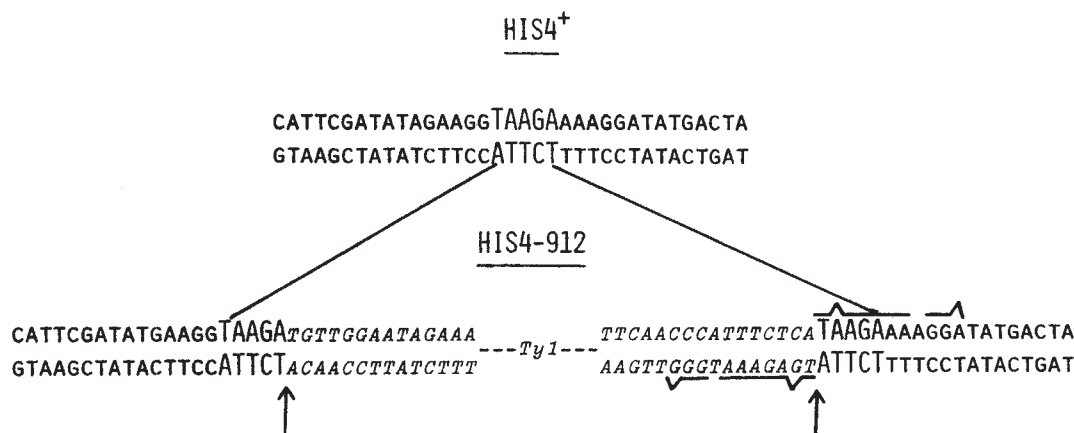


Fig. 5 The sequences of *His4⁺* and *his4-912* junctions. The arrows are at the junction points. The regions homologous to *His4⁺* in the sequences of the two junctions are shown in roman lettering, the *Ty1* element in italics; the 5-base pair duplication is indicated by large letters. The lines in either direction from the right junction demonstrate the inverted repeat; a (-) indicates identity and a (^) indicates purine-purine partial identity.

Ty1 is inserted into the region of the *his4* promoter

A comparison of the sequences of *His4⁺* and *his4-912* demonstrates that in the *his4-912* mutation *Ty1* lies in the 5'-non-coding region, 161 base pairs from the ATG responsible for initiation of *his4* translation (Fig. 2). Because the sequence encoding the *his4* polypeptide is unaffected and the mutation is distant from the structural gene, the *His⁻* phenotype of *his4-912* is likely to result from a block in the expression of the gene at the level of transcription.

Unfortunately, we do not know exactly where *his4* transcription begins. Comparison of the *his4* sequence with that of known eukaryotic promoters limits the position possible for *his4* transcriptional initiation. Regions where eukaryotic RNA polymerase II initiates have two common features. Translation begins at the first AUG following the site of transcriptional initiation²¹. As shown by studies on the *cycl* gene of *S. cerevisiae*, the only signal required for initiation of translation is the AUG²². This observation is supported by comparisons of 5'-non-coding regions of many eukaryotic mRNAs²³. Based on these considerations the *his4* message must initiate downstream from the first ATG in the non-coding region (-88) between -87 and +1.

The second common feature of eukaryotic promoters is a sequence which is thought to be a recognition site for RNA polymerase II. Nearly all promoters have a sequence similar to 5'-TATAAATA-3' (the 'Hogness box') which is centred approximately 28 base pairs on the 5' side of the initial nucleotide of the message²⁴. The promoter regions of the yeast genes *his3*²⁵, *cycl*²⁶, and the gene encoding glyceraldehyde-3-phosphate dehydrogenase²⁷ all include a sequence related to the Hogness box. If *his4* mRNA begins between -87 and +1, any Hogness box must lie between nucleotides -120 and -25. Two sequences similar to the Hogness box occur in this region: 5'-TATATAATA-3', centred on -119, and 5'-CATAAAAA-3', centred on -67. The *His⁻* phenotype of *his4-912* does not result from disruption of the DNA sequence in the region of either Hogness box.

Preliminary RNA blotting experiments indicate that the normal *his4* message is not produced in *his4-912* (C. Ilgen, personal communication). One possible explanation for the *His4⁻* phenotype of the *his4-912* mutation is that transcription initiated within the *Ty1* element disrupts the initiation of transcription at the normal *his4* promoter. The other possibility is that *Ty1* inserted into a site required for polymerase recognition. This would mean that the promoter includes sites in the region of -161 as well as the Hogness box.

The *his4-912R1* revertant created by excision of most of the *Ty1* element can distinguish between these two possibilities. In *his4-912R1* there is no new promoter either at the *his4*-proximal junction of the insertion or in the insertion itself, as this region is identical to that of *his4-912*. A promoter created at the *his4*-distal junction would result in transcription of the element

itself. The element includes nine ATG codons each followed by several in-frame nonsense codons before the beginning of *his4*. Such a transcript would fail to express *His4⁺* because translation would initiate and terminate before the *his4* gene. The simplest hypothesis is that the *his4* promoter is located between the *his4-912* site (-161) and the beginning of the gene. The *His⁻* phenotype of *his4-912* could be attributed to transcriptional competition. Deletion of most of the element by excision eliminates the *Ty1* transcript so that initiation at the *his4* promoter occurs normally. The residual cold sensitivity of *his4-912* is not explained by this model. This phenotype could be due to some other pleiotropic effect of the δ element (such as that due to a minor transcript), or the presence of a δ element could alter the *his4* promoter. Many predictions of this model are being tested by RNA blotting experiments.

Sequence analysis of the *Ty1* element from *his4-912* demonstrates that *Ty1* is a transposon and is capable of transposition by a mechanism similar to that of bacterial transposons. Southern blotting analysis of *His⁺* revertants of *his4-912*⁵ suggest that the element is also capable of promoting chromosomal aberrations similar to those catalysed by bacterial transposons. One mutant may be the result of a transposition-mediated inversion, similar to those seen with bacteriophage Mu²⁸. The *Ty1* element also promotes events which differ from any observed in bacteria. These include deletions ending at apparently random locations in the element, chromosomal aberrations with breakpoints outside the element, and 'concerted' events^{6,7}, and suggest that the mechanism of action of *Ty1* is significantly different from that of prokaryotic transposons.

We thank Deborah Chaleff, Thomas Donahue, John Lis and Shirleen Roeder for their critical reading of this manuscript and for helpful discussions. This work was supported by NIH grant GM15408. P.J.F. was an American Cancer Society Postdoctoral Fellow. Experiments involving recombinant DNA techniques were carried out according to the NIH guidelines.

Received 25 February; accepted 16 May 1980.

- Kleckner, N. *Cell* **11**, 11-23 (1977).
- Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (eds) in *DNA Insertion Elements, Plasmids and Episomes*, 3-11 (Cold Spring Harbor Laboratory, New York, 1977).
- McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **21**, 197-216 (1956).
- Potter, S. S., Brorein, W. J., Dunsin, P. & Rubin, G. M. *Cell* **17**, 415-427 (1979).
- Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16**, 739-751 (1979).
- Roeder, G. S. & Fink, G. R. *Cell* (in the press).
- Chaleff, D. T. & Fink, G. R. *Cell* (in the press).
- Hinnen, A., Farabaugh, P., Ilgen, C. & Fink, G. R. in *Eukaryotic Gene Regulation* (eds Axel, R., Maniatis, T. & Fox, C. F.) 43-50 (Academic, New York, 1979).
- Farabaugh, P. & Fink, G. R. *Gene* (in the press).
- Kessey, J. K., Bigelis, R. & Fink, G. R. *J. biol. Chem.* **254**, 7427-7433 (1979).
- Grindley, N. D. F. & Sherrat, D. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1257-1261 (1978).
- Kopecko, D. J., Brevet, J. & Cohen, S. N. *J. molec. Biol.* **108**, 333-360 (1976).
- Barth, P. T. & Grintner, N. J. *J. molec. Biol.* **113**, 455-474 (1977).
- Johnsrud, L., Calos, M. & Miller, J. H. *Cell* **13**, 411-418 (1978).
- Botstein, D. & Kleckner, N. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 185-203 (Cold Spring Harbor Laboratory, New York, 1977).
- Foster, T. J. *Molec. gen. Genet.* **154**, 305-309 (1977).
- Grinstead, J., Bennett, P. M., Higginson, S. & Richmond, M. H. *Molec. gen. Genet.* **166**, 313-320 (1978).
- Kleckner, N. *Cell* **16**, 711-720 (1979).

19. Tu, C.-P. D. & Cohen, S. N. *Cell* **19**, 151-160 (1980).
20. MacHattie, L. A. & Jackowski, J. B. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 219-228 (Cold Spring Harbor Laboratory, New York, 1977).
21. Kozak, M. *Cell* **15**, 1109-1123 (1979).
22. Sherman, F., Stewart, J. W. & Schweingruber, A. M. *Cell* (in the press).
23. Baralle, F. E. & Brownlee, G. G. *Nature* **274**, 84-87 (1978).
24. Gannon, F. et al. *Nature* **278**, 428-433 (1979).
25. Struhl, K. & Davis, R. W. *Cell* (in the press).
26. Smith, M. et al. *Cell* **16**, 753-761 (1979).
27. Holland, J. P. & Holland, M. J. *J. biol. Chem.* **254**, 9839-9845 (1979).
28. Faellen, M. & Toussaint, A. J. *Bact.* **136**, 477-483 (1978).
29. Smith, D. & Calvo, J. *Nucleic Acid Res.* (in the press).
30. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977); *Meth. Enzym.* **65** (in the press).

A signal sequence is not sufficient to lead β -galactosidase out of the cytoplasm

Felipe Moreno^{*§}, Audree V. Fowler[†], Michael Hall[‡], Thomas J. Silhavy[‡],
Irving Zabin[†] & Maxime Schwartz^{*}

^{*}Unité de Génétique Moléculaire, Institut Pasteur 25 rue du Dr Roux, 75724 Paris Cedex 15, France

[†]Department of Biological Chemistry, UCLA School of Medicine and Molecular Biology Institute, University of California, Los Angeles, California 90024

[‡]Frederick Cancer Research Center, Cancer Biology Program, Frederick, Maryland 21701

Escherichia coli strains have been constructed in which *lacZ*, the gene for the cytoplasmic enzyme β -galactosidase, is fused to *lamB*, the gene for an outer membrane protein. One such strain produces a β -galactosidase which remains cytoplasmic even though it possesses the complete signal sequence of the *lamB* protein precursor at the amino-terminal end.

ACCORDING to the signal hypothesis, proteins secreted from eukaryotic cells are synthesized as precursors with an extra NH₂-terminal sequence of predominantly hydrophobic amino acids^{1,2}. That 'signal sequence' initiates binding of the translation complex to the endoplasmic reticulum, facilitating transfer of the protein across the membrane concomitant with translation. In most cases the signal sequence is removed by a 'signal peptidase' from the growing polypeptide during transfer, giving rise to the final protein product, which is localized on the luminal side of the endoplasmic reticulum. Evidence for this scheme in animal cells has come from the isolation, *in vitro*, of numerous precursors of secreted proteins, the discovery that they have extra amino acid sequences (15-30 residues) at their NH₂-terminus and the demonstration that secreted proteins are synthesized on polysomes bound to the endoplasmic reticulum whereas cytoplasmic proteins are synthesized on free polysomes³⁻⁷. On the basis of similar evidence, the hypothesis was extended to include the localization of non-cytoplasmic proteins in bacteria, with the cytoplasmic membrane having a role analogous to that of the eukaryotic endoplasmic reticulum⁸⁻¹³. (By non-cytoplasmic we mean proteins of the cytoplasmic or outer membranes, of the periplasmic space, or truly secreted proteins.)

We are using a genetic approach to evaluate the importance of the signal sequence in the export process¹⁴⁻¹⁸. One non-cytoplasmic protein which we have been studying is the product of gene *lamB* in *Escherichia coli*. Located in the outer membrane, it is involved in the transport of maltose and maltodextrins, and also serves as a receptor for several bacteriophages, including λ (refs 19-22). Comparison of the NH₂-terminal sequence of the mature protein²³ with that of the primary gene product^{24,25} showed that *lamB* protein is synthesized as a precursor containing 24 extra amino acids at the NH₂ terminus. Mutations affecting this extra sequence prevent the export of the protein, which remains in the cytoplasm as an uncleaved precursor²⁶. From this we conclude, as predicted by the signal hypothesis, that the extra sequence in the precursor of non-cytoplasmic proteins is essential for their export.

The question then is whether such an extra sequence possesses all the information required to lead any polypeptide out of the cytoplasm.

By means of gene fusion we have already shown that a polypeptide coded by a *lamB-lacZ* hybrid gene was exported, to some extent, to the outer membrane, although β -galactosidase, the product of *lacZ*, is normally cytoplasmic¹⁵. The amount of *lamB* DNA present in this hybrid gene (called 42-1) was 200-300 codons (out of a total of about 500 for the entire gene). We now describe two other fusion strains, in which there is much less *lamB* DNA in the hybrid gene. Our results demonstrate that the presence of a signal sequence at the NH₂ terminus of β -galactosidase does not suffice to lead this polypeptide out of the cytoplasm.

Construction and characterization of fusion strains

The application of gene fusion *in vivo*²⁷ to the selection of strains with *lamB-lacZ* hybrid genes has been described previously¹⁵ (Fig. 1). It involves the insertion of phage Mu DNA into the *lamB* gene, the integration of a λ plac prophage into the Mu prophage, and a deletion of the Mu prophage. Two fusions will be described here, 61-4 and 52-4, derived from Mu insertions 61

Table 1 Purification of the hybrid proteins

	61-4		52-4	
	Total activity (units)	Specific activity (units per mg protein)	Total activity (units)	Specific activity (units per mg protein)
French press extract	9.9 × 10 ⁸	2.5 × 10 ⁴	3.3 × 10 ⁹	5.1 × 10 ⁴
38% Ammonium sulphate precipitate	4.3 × 10 ⁸	7.6 × 10 ⁴	1.8 × 10 ⁹	2.2 × 10 ⁵
DEAE-Sephadex A50 eluate	3.0 × 10 ⁸	5.5 × 10 ⁵	1.2 × 10 ⁹	4.4 × 10 ⁵

The 61-4 and 52-4 hybrid proteins were purified from 700 g and 1,000 g, respectively, of cells grown to late exponential phase at 30°C in M63 minimal medium²⁴ supplemented with 0.8% maltose. The purification was essentially as described by Miller³⁴ except that: (1) the extraction was carried out in the presence of 30 μ g ml⁻¹ *p*-toluylsulphonyl fluoride to inhibit proteases present in the extract, (2) a precipitation step with 3% streptomycin sulphate was carried out before the ammonium sulphate precipitation, (3) DEAE Sephadex A50 was used instead of DEAE cellulose, (4) the sucrose gradient step was omitted and (5) 10⁻⁴ M Na-EDTA and 10⁻¹ M β -mercaptoethanol were present in all buffers. The purified proteins were kept in the form of a 50% ammonium sulphate precipitate.

§ Present address: Servicio de Microbiología, Centro Especial Ramon y Cajal, Madrid-34, Spain.

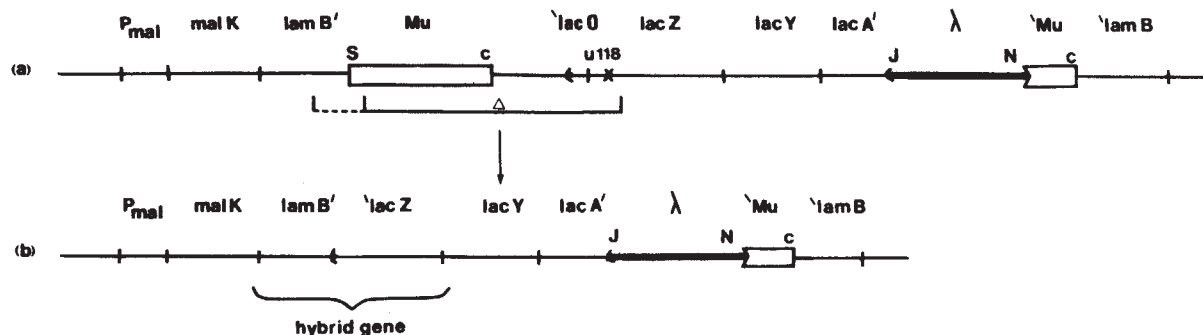


Fig. 1 Scheme for isolating *lamB-lacZ* fusion strains. Because the whole procedure has been previously described¹⁵, only the last step is shown here. A thermoinducible Mu(cts) prophage was first inserted into *lamB* and a λ lac Mu prophage was then inserted into the Mu prophage, by recombination between homologous Mu DNA sequences. The structure of the resulting strain is shown in *a*. Pmal is the promoter of the *malK-lamB* operon. The S- and c-ends of the Mu prophage⁴⁴ are indicated. Primes, as in *lamB'* or *lacO'*, indicate that the genetic region referred to is deleted on the side the prime is written. U118 is an ochre mutation located in codon 17 of gene *lacZ* (ref. 45). Deletion events (Δ) removing all or most of the Mu(cts) prophage and extending past the U118 mutation yield strains whose genetic structure are shown in *b*. Such strains can be selected as thermoresistant (loss of the thermoinducible Mu(cts) prophage) and Lac⁺ (elimination of U118 and production of a hybrid protein which has β -galactosidase activity).

and 52, respectively (Fig. 2). The presence of *lamB* DNA within the 52-4 hybrid gene was demonstrated by deletion mapping^{15,28,29}. Because fusion 61-4 was derived from a Mu insertion located at the very beginning of *lamB*, the presence of *lamB* DNA in this hybrid gene could not be ascertained by this technique. Accordingly, we had to insure that *lacZ* was not fused to the promoter proximal gene *malK*. This was accomplished by demonstrating (as described in ref. 30) that the introduction of a non-polar amber mutation in *malK* (*malK*3 or *malK*4), in the *cis* position with respect to these fusions, had no effect on the synthesis of β -galactosidase. The introduction of such a mutation into known *malK-lacZ* fusions leads to the disappearance of β -galactosidase activity.

To locate the hybrid proteins in the cell, we prepared extracts and separated the cytoplasmic and envelope fractions as described elsewhere¹⁴⁻¹⁶. The hybrid proteins can be detected either by their enzymatic activity or by their migration on polyacrylamide gels in the presence of SDS. About 10% of the β -galactosidase activity sediments with the envelope in the case of 61-4, and 15% with 52-4. However, these figures, which result from assays carried out on unwashed envelope pellets, certainly represent overestimations of the amount of hybrid protein which is truly membrane bound. Indeed, when extracts of wild-type strains are centrifuged in the same conditions, about 10% of authentic β -galactosidase also sediments with the envelopes. The gel pattern in Fig. 3 confirms that the hybrid proteins are mostly cytoplasmic: the amount of envelope-bound protein is undetectable with 61-4 and is less than 3% with 52-4. Figure 3 also shows that the 61-4 polypeptide is slightly shorter than 52-4.

When 61-4 or 52-4 cells are submitted to cold osmotic shock³¹, less than 2% of β -galactosidase activity, but about 90% of maltose binding protein³², are released into the medium. Thus, the two hybrid proteins are not in the periplasmic space.

According to our fractionation data, therefore, the 61-4 and 52-4 hybrid proteins are chiefly in the cytoplasm, perhaps a very small fraction of the latter being bound to the membrane. That location accords with the growth properties of these two strains, which grow normally in the presence of maltose, an inducer of the *malK-lamB* operon, whereas other *lamB-lacZ* strains (such as 42-1, Figs 2, 3), which export some of the hybrid protein to the outer membrane, are sensitive to maltose^{15,16}.

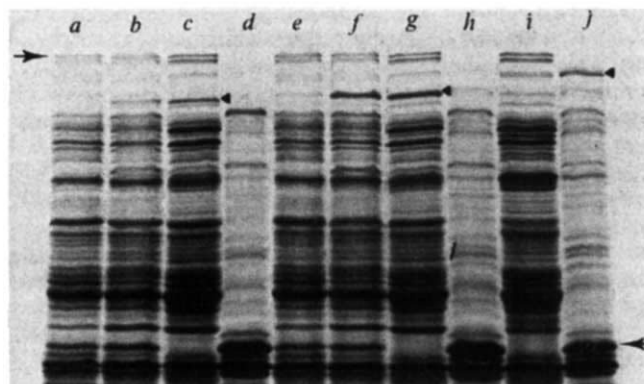


Fig. 3 SDS-polyacrylamide gel electrophoresis of whole cell extracts and cellular fractions of *lamB-lacZ* fusion strains. Whole cell extracts were prepared by growing the cells in the indicated media to mid-logarithmic phase, resuspending in 0.1 volume SDS sample buffer⁴⁶ and boiling for 2 min. Cytoplasmic and membrane fractions were prepared as previously described¹⁴ by centrifuging at 100,000g for 1 h extracts obtained by sonication. All 'induced' samples correspond to cells grown at 30 °C for 4 h in maltose-glycerol M63 minimal medium supplemented with 5% complete medium (LB). 'Uninduced' samples correspond to cells grown in glucose M63 medium. The acrylamide concentration was 9%. Channels *a-d* are various preparations from the 61-4 fusion strain, channels *e-h* correspond to the 52-4 strain and channels *i* and *j* to the 42-1 strain. *a* and *e* are uninduced whole cell extracts; *b* and *f* are induced whole cell extracts; *c*, *g* and *i* are induced cytoplasmic fractions; *d*, *h* and *j* are induced membrane fractions. The membrane fractions displayed here correspond to 7.5 times more cells than the cytoplasmic fractions. The solid arrowheads indicate the *lamB-lacZ* hybrid proteins. The complete arrows indicate cytoplasmic and membrane markers: at the upper left the β and β' subunits of RNA polymerase and, at the lower right, major outer membrane protein Ia.

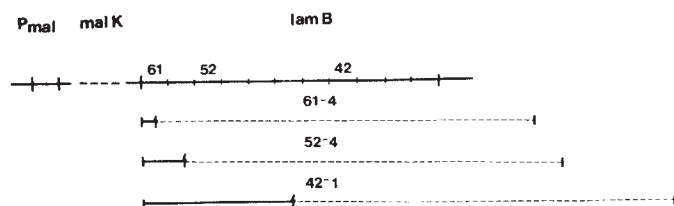


Fig. 2 Genetic structure of the *lamB-lacZ* fusions. Previous work²⁸ allowed us to divide gene *lamB* into 11 deletion intervals. The position of Mu insertions 61, 52 and 42 is shown on the first line. The structure of fusions derived from these Mu insertions is indicated below the map, with the *lamB* part of the hybrid gene shown as a heavy line and the *lacZ* part as an interrupted line. Only fusions 61-4 and 52-4 are described in this article. fusion 42-1 is described elsewhere¹⁵; it produces a hybrid protein which is partly exported to the outer membrane (see Fig. 3).

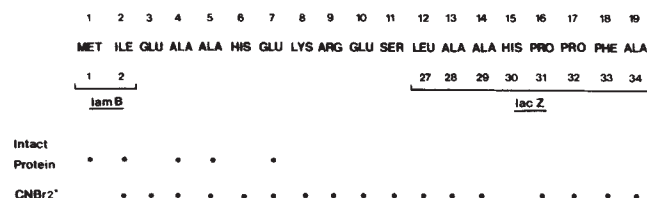


Fig. 4 Amino-terminal sequence of the 61-4 *lamB-lacZ* hybrid protein. All sequencing was done using 0.1 M Quadrol in a Beckman 890C sequencer. The CNBr²* fragment was sequenced using a dual wash program⁴⁷. The intact carboxymethylated protein was sequenced with a program similar to that of Hunkapiller and Hood⁴⁸. All other procedures were as described previously³⁶ except that, recently, identification has also used HPLC. ● Indicates residue identified. The numbers above the sequence correspond to the position of the residue in the 61-4 protein, the numbers below correspond to the position of the residues in the *lamB* and *lacZ* gene products. According to the DNA sequencing data the *lamB* gene starts on the mRNA with two successive AUG codons. Therefore, the *lamB* protein precursor might be expected to start with two successive methionine residues. However the two hybrid proteins described here start with a single methionine residue, and there is some evidence that this is also true of the *lamB* protein precursor synthesized *in vitro* (ref. 25). We do not know whether this is due to an initiation of translation at the second, rather than the first AUG, or to a proteolytic cleavage of the first methionine. This introduces an ambiguity in the numbering of residues in the *lamB* gene product. We decided to call residue number 1 the single methionine which could be detected at the NH₂-terminal end of the hybrid proteins.

The 61-4 and 52-4 hybrid genes were transduced into other strains of *E. coli* K12 (see below). The transductants grew normally in the presence of maltose and synthesized a cytoplasmic β -galactosidase. Thus, the original fusion strains carried no unlinked mutation which might have prevented the export of hybrid protein.

Purification of the hybrid proteins

To facilitate purification we used genetic techniques to increase the amount of hybrid protein synthesized by fusion strains. Any hybrid gene obtained by our technique of gene fusion is located close to a λ prophage (Fig. 1). Induction of fusion strains with UV light yields λ transducing phages which carry the hybrid gene. When such phages are used to lysogenize another strain, multiple lysogens are obtained at a very high frequency, and these can be recognized by the amount of β -galactosidase which

they synthesize³³. After infecting MC4100 cells (deleted for the *lac* operon²⁷) with a λ 61-4 or a λ 52-4 transducing phage, we selected presumed triple lysogens, which synthesized approximately three times as much β -galactosidase as the original fusion strain. Although they were of limited stability, these polylysogens could be used for protein purification, provided the number of transfers after their isolation was minimized.

Hybrid proteins were purified essentially according to the scheme used for authentic β -galactosidase³⁴ (Table 1). The 52-4 protein had a stability comparable to that of authentic β -galactosidase. The 61-4 protein, on the other hand, was rather unstable. For instance, it lost 90% of its activity after 10 days at 4°C when kept in solution after elution from the DEAE Sephadex column.

Determination of NH₂ terminus of hybrid proteins

Hybrid protein sequences were analysed by techniques developed for use on a hybrid protein containing β -galactosidase and *lac* repressor sequences³⁵. In wild-type β -galactosidase, the first two methionine residues are at positions 2 and 92 (ref. 36), and residues 3–92 yield a cyanogen bromide fragment called CNBr2. Hybrid proteins obtained as described in Fig. 1 legend always lack at least the first 17 amino acids of wild-type β -galactosidase, eliminating the first methionine. If the amino acid sequence which replaces the normal β -galactosidase NH₂ terminus contains no internal methionine, only one new fragment (CNBr2*) should result from treatment of the hybrid protein with cyanogen bromide. Indeed, with the 61-4 hybrid protein a single additional cyanogen bromide fragment was found that reacted with antibodies directed against wild-type CNBr2. The NH₂-terminal sequence of this peptide is the same as that of the intact hybrid protein (Fig. 4). The 52-4 hybrid protein, on the other hand, yielded three new peptides, due to the presence of methionine residues at positions 18 and 23 in the *lamB* portion of the hybrid protein²⁵. To establish the sequence of the 52-4 hybrid protein up to the β -galactosidase sequence, it was therefore necessary to analyse the NH₂-terminal sequence on the intact protein, to purify the CNBr2* fragment and sequence its NH₂-terminal region, and to purify and sequence the smallest of the additional cyanogen bromide fragments.

As Figs 4 and 5 show, the 61-4 and 52-4 proteins contain *lacZ*

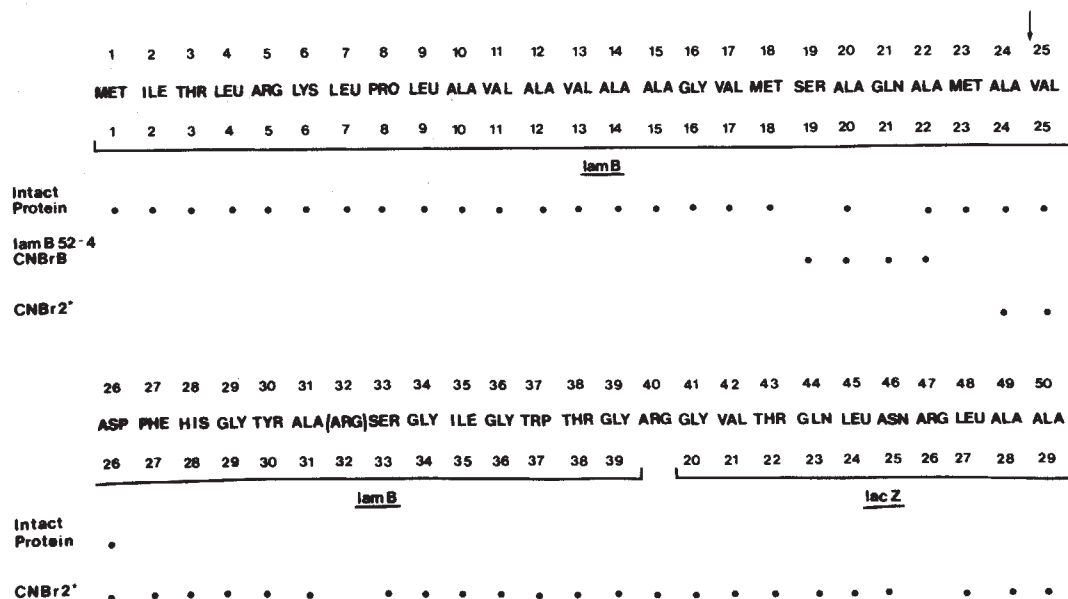


Fig. 5 Amino-terminal sequence of the 52-4 *lamB-lacZ* hybrid protein. See Fig. 3 legend. The arrow indicates the site of cleavage by the signal peptidase. *lamB* 52-4 CNBrB is one of the three cyanogen bromide fragments obtained from the 52-4 protein but not from authentic β -galactosidase.

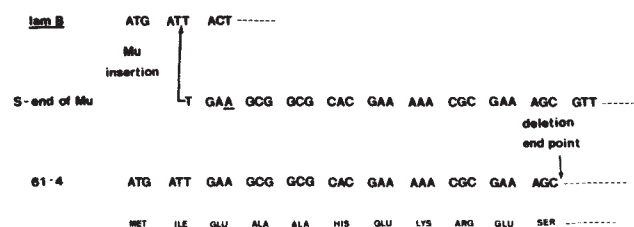


Fig. 6 A reconstruction of the events which yielded the 61-4 fusion. The fragment of *lamB* sequence on the first line is from ref. 25. The Mu sequence on the second line is from Kahman and Kamp³⁹ and Allet⁴⁰. (The sequence published by Allet⁴¹ differs in two respects from that presented here: (1) the "A" underlined here is missing and (2) the prophage sequence would start at the "G" immediately following this "A".) Only the 'sense' strand of the DNA is shown here. The NH₂-terminal sequence of the 61-4 protein (shown on the last line) can be explained if we assume that the original Mu insertion and the final deletion (see Fig. 1) occurred as shown here by the two arrows.

polypeptides in which the first 26 or 19 amino acids, respectively, have been replaced by 11 or 40 other amino acids. The 61-4 polypeptide is therefore 15 amino acids shorter than the authentic *lacZ* polypeptide, whereas the 52-4 polypeptide is 21 amino acids longer. The fact that these sequences could be determined unambiguously shows that there was no extensive site-specific cleavage in the NH₂-terminal region of the hybrid polypeptides (see also Fig. 3).

Discussion

Because the NH₂-terminal sequence of the 61-4 hybrid protein contains only the first two amino acids of the precursor of the *lamB* protein (Fig. 4), it is not surprising that the hybrid protein remains in the cytoplasm. Indeed, other experiments have shown that most or all of the 24-amino acid signal sequence is essential for export of *lamB* protein from the cytoplasm^{16,26,29}. This is also consistent with the observations that *phoA-lacZ* and *malE-lacZ* hybrid proteins containing truncated signal sequences remain in the cytoplasm, whereas wild-type *phoA* and *malE* products are exported to the periplasm^{37,38}. The amino acid sequence also revealed that the deletion generating the fusion removed the first 26 codons of the *lacZ* gene. This is the largest amount found missing in any fusion strain synthesiz-

ing a hybrid protein still endowed with β -galactosidase activity^{35,37}.

One unexpected feature of the 61-4 protein is that its amino acid sequence in residues 3-11 is found neither at the NH₂-terminal end of the *lamB* protein precursor nor at that of β -galactosidase. This sequence is actually that which would be coded by the DNA sequence located at the S-end of prophage Mu^{39,40}. Therefore, we believe that, in the case of the 61-4 strain, the deletion which constituted the final event in the creation of the fusion (Fig. 1) failed to remove a small portion of prophage Mu (Fig. 6). (There is some controversy regarding the nucleotide sequence in Mu DNA; our results could not be explained using the sequence originally reported by Allet⁴¹).

As might have been predicted from genetic data (Figs 1, 2), the 52-4 protein does not contain Mu-coded amino acids (Fig. 5). From residue 1 to 39 its sequence corresponds to the NH₂ terminus of the *lamB* protein precursor. Except for residue 40, which corresponds to the fusion joint, the rest of the sequence is from β -galactosidase. Therefore, this protein contains the entire signal sequence of the *lamB* protein precursor and is, nevertheless, chiefly located in the cytoplasm (Fig. 3). One interpretation of this result would be that the signal sequence is not sufficient information to initiate the transfer of a polypeptide through the cytoplasmic membrane. If the additional information required were encoded over a large fraction of the polypeptide, alternatives⁴² to the signal hypothesis would have to be considered. Another possibility is that the signal sequence does suffice to initiate the transfer but that severe restrictions exist as to which amino acid sequences can be extruded through a membrane⁴³. The enzyme β -galactosidase, which is normally cytoplasmic, may contain unfavourable sequences¹⁷. These sequences might block the transfer of the 52-4 protein at an early stage, so that this protein, weakly bound to the membrane by its NH₂-terminus, may fall back into the cytoplasm as synthesis proceeds. Whatever the interpretation, the fact remains that adding a signal sequence to at least one cytoplasmic protein— β -galactosidase—is not a sufficient condition to alter to any significant extent the cellular location of this protein.

We thank M. Hours for growing the cells necessary for protein purification. This work was supported by grants from the Délégation Générale à la Recherche Scientifique et Technique (77.7.1294) and the Centre National de la Recherche Scientifique (LA 270) to M.S., from the NSF (78-1-9974) and USPHS (AI 04181) to I.Z., and from the National Institute of General Medical Science (GM25524) to T.S.

Received 23 November 1979; accepted 23 April 1980.

- Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. *Nature new Biol.* **239**, 117-120 (1972).
- Blöbel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835-851 (1975).
- Palade, G. *Science* **189**, 347-358 (1975).
- Rothman, J. E. & Lodish, H. F. *Nature* **269**, 775-780 (1977).
- Kreibich, G., Ulrich, B. L. & Sabatini, D. D. *J. Cell Biol.* **77**, 464-487 (1978).
- Austen, B. M. *FEBS Lett.* **103**, 308-313 (1979).
- Blöbel, G. *et al. Symp. Soc. exp. Biol.* **33**, 9-36 (1979).
- Randall, L. & Hardy, S. *Eur. J. Biochem.* **75**, 43-53 (1977).
- Smith, W. P., Tai, P. C., Thompson, R. C. & Davis, B. D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2830-2834 (1977).
- Inouye, H. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1440-1444 (1977).
- Di Rienzo, J. M., Nakamura, K. & Inouye, M. A. *Rev. Biochem.* **47**, 481-532 (1978).
- Lin, J. J. C., Kanazawa, H., Ozols, J. & Wu, H. C. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4891-4895 (1978).
- Chang, C. N., Model, P. & Blöbel, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1251-1255 (1979).
- Silhavy, T. J., Casadaban, M. J., Shuman, H. A. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3423-3427 (1976).
- Silhavy, T. J., Shuman, H. A., Beckwith, J. & Schwartz, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5411-5415 (1977).
- Emr, S. D., Schwartz, M. & Silhavy, T. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5802-5806 (1978).
- Bassford, P. J. Jr, Silhavy, T. J. & Beckwith, J. R. *J. Bact.* **139**, 19-31 (1979).
- Silhavy, T. J., Bassford, P. J. Jr & Beckwith, J. R. in *Bacterial Outer Membranes* (ed. Inouye, M.) 203-254 (Wiley, New York, 1979).
- Randall-Hazelbauer, L. & Schwartz, M. *J. Bact.* **116**, 1436-1446 (1973).
- Wandersman, C. & Schwartz, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5636-5639 (1978).
- Wandersman, C., Schwartz, M. & Ferenci, T. *J. Bact.* **140**, 1-13 (1979).
- Roa, M. *J. Bact.* **140**, 680-686 (1979).
- Endermann, R., Hindennach, I. & Henning, U. *FEBS Lett.* **88**, 71-74 (1978).
- Randall, L. L., Hardy, S. J. S. & Josefsson, L. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1209-1212 (1978).
- Hedgpeth, J., Clément, J. M., Marchal, C., Perrin, D. & Hofnung, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2621-2625 (1980).
- Emr, S., Hedgpeth, J., Clément, J. M., Silhavy, T. J. & Hofnung, M. *Nature* **285**, 82-85 (1980).
- Casadaban, M. *J. molec. Biol.* **104**, 541-555 (1976).
- Raibaud, O., Roa, M., Braun-Breton, C. & Schwartz, M. *Molec. gen. Genet.* **174**, 241-248 (1979).
- Emr, S. & Silhavy, T. J. *J. molec. Biol.* (in the press).
- Débarbouillé, M. & Schwartz, M. *J. molec. Biol.* **132**, 521-534 (1979).
- Neu, H. C. & Heppel, L. A. *J. biol. Chem.* **240**, 3685-3692 (1965).
- Kellermann, O. & Szmecman, S. *Eur. J. Biochem.* **47**, 139-149 (1974).
- Sanzey, B. *J. Bact.* **138**, 40-47 (1979).
- Miller, J. H. in *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
- Brake, A. J., Fowler, A. V., Zabin, I., Kania, J. & Muller-Hill, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4824-4827 (1978).
- Fowler, A. V. & Zabin, I. *J. biol. Chem.* **253**, 5521-5525 (1978).
- Sarthy, A., Fowler, A., Zabin, I. & Beckwith, J. *J. Bact.* **139**, 932-939 (1979).
- Bedouelle, H. *et al. Nature* **285**, 78-81 (1980).
- Kahman, R. & Kamp, D. *Nature* **280**, 247-250 (1979).
- Allet, B. *Cell* **16**, 123-129 (1979).
- Allet, B. *Nature* **274**, 553-558 (1978).
- Wickner, W. A. *Rev. Biochem.* **48**, 23-45 (1979).
- Van Heijne, G. & Blomberg, C. *Eur. J. Biochem.* **97**, 175-181 (1979).
- Bukhari, A. I. A. *Rev. Genet.* **10**, 389-412 (1976).
- Zabin, I., Fowler, A. V. & Beckwith, J. R. *J. Bact.* **133**, 437-438 (1977).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Brauer, A. W., Margolis, M. N. & Haber, E. *Biochemistry* **14**, 3029-3035 (1975).
- Hunkapiller, M. W. & Hood, L. E. *Biochemistry* **17**, 2124-2144 (1978).

LETTERS

Spontaneous production of cosmological fluctuations

Craig J. Hogan

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

In the 'standard' hot big bang, fluctuations which give rise to galaxies and galaxy clusters must be present in the initial conditions; that is, they are already present at the earliest epoch when general relativity is valid. In a structureless big bang there seems to be only one way that the necessary fluctuations can be explained within the scope of conventional physics: a uniform universe with low specific entropy ($S \leq 1$) can spontaneously generate fluctuations by shattering into macroscopic fragments. Such a critical point probably occurs at nuclear density during the transition from pion condensate to nucleon gas; shattering at this epoch would create fragments of the order of a Chandrasekhar mass, providing both the lumpiness necessary to produce galaxy clusters and a potential energy source for the microwave background radiation.

The universe seems to be uniform on the very largest scales, and we know that it is fundamentally non-uniform on the smallest scales. Furthermore, we know that small irregularities introduced into a homogeneous Friedmann universe can grow with time. The obvious temptation is to try to show that all inhomogeneity, even on astronomical scales, ultimately originates from quantum effects operating in a universe which is initially as uniform as possible. Proposals along these lines have been made using, for example, Heisenberg uncertainty in the metric¹ and thermal fluctuations in particle positions². Several authors have also invoked various types of phase transitions, including vacuum phase transitions at very early times^{3,4}, a nucleated quark-nucleon phase transition⁵, and the conventional solid- or liquid-gas transitions^{6,7}. Critical-point transitions are particularly good for generating fluctuations spontaneously because they afford an enormous amplification of quantum discreteness which is in principle limited only by conditions of causality: if such a transition occurs, matter is unstable and must shatter into macroscopic fragments^{6,7}. A uniform expansion can 'tear apart' a uniform medium to leave a highly non-uniform expanding 'gas' of macroscopic lumps. Some fundamental constraints on such schemes are described here and a model proposed which does not rely on a speculative departure¹⁻³ from classical general relativity.

Although some quantum inhomogeneity always exists, it is possible to show that in the standard hot big bang no classical process can create the actual observed scale of inhomogeneity. Let us assume that quantum effects operating at epoch t_1 introduce density fluctuations with amplitude $\delta\rho/\rho = \delta_M$ on a co-moving scale labelled by its baryon charge M . At the present time t_0 , ignoring dissipation effects and assuming $\Omega = 1$, linear Lifshitz gravitational growth gives

$$\delta_M(t_0) = (a_1/a_0)^2 (t_0/t_1)^2 \delta_M(t_1) \quad (1)$$

independent of the equation of state, where a is the scale factor. If gravitating energy and momentum are locally conserved in the sense required by classical general relativity⁹ (which is generally true even during vacuum phase transitions^{3,4}), then the spectrum of growing fluctuations introduced by any irreversible process on scales larger than the horizon must be^{8,10-12,32}

$$\delta_M(t_1) \propto M^{-7/6} \quad (2)$$

This spectrum has been shown^{10,12,32} to arise in a perfect fluid with an arbitrarily variable equation of state; any conservative matter is gravitationally identical to such a fluid on scales sufficiently large that the off-diagonal components of $T^{\mu\nu}$ are

uncorrelated. Moreover, causal processes cannot move matter further than ct_1 , so the density contrast on the horizon scale must be less than unity:

$$(\delta\rho/\rho)_{ct_1} \equiv \delta_H(t_1) < 1 \quad (3)$$

Using equation (2), causality limits the amplitude of $(\delta\rho/\rho)$ on any mass scale:

$$\delta_M(t_1) < (M/M_H(t_1))^{-7/6} \quad (4)$$

where M_H is the baryon charge within the horizon at t :

$$M_H(t_1) = M_H(t_0)(a_1/a_0)^{-3}(t_1/t_0)^3 \quad (5)$$

Thus in an initially uniform universe

$$\delta_M(t_0) < (a_1/a_0)^{-3/2}(t_1/t_0)^{3/2}(M/M_H(t_0))^{-7/6} \quad (6)$$

The actual universe is observed⁸ to have $\delta_M = 1$ on a length-scale $r_* \approx 3h^{-1}$ Mpc (where $h \equiv H_0/100$); since $ct_0 = 3,000 h^{-1}$ Mpc, we need $\delta_{M*} = 1$ for

$$M_*/M_H = 10^{-9} \quad (7)$$

This requirement, the Friedmann relations between a and t , and equation (6) put a lower limit on (a_1/a_0) , that is an upper limit $z_{\max}$ on the redshift at which fluctuations can be created by a causal process. The reason for this limit is that with an initial spectrum of the form of equation (2) the amplitude $(\delta\rho/\rho)_H$ on the horizon scale (or any constant fraction thereof) decreases with time; although the amplitude of fluctuations increases, the universe actually grows 'smoother' with time. No classical process at a redshift greater than $z_{\max}$ can generate fluctuations which give rise to the observed structure.

Now in a hot universe¹³

$$(a_1/a_0) \equiv (t_1/t_{eq})^{1/2}(t_{eq}/t_0)^{2/3} \quad (8)$$

where the epoch of equal matter and radiation densities is

$$(t_{eq}/t_0) = 10^{-7} (S/10^8)^2 \quad (9)$$

where S denotes the specific entropy in the fireball phase (again assuming $\Omega h^2 = 1$). The maximum redshift at which fluctuations can form if $\Omega h^2 \leq 1$ is

$$z_{\max} < 10^9 (S/10^8)^{-2/3} \quad (10)$$

corresponding to a total (radiation + matter) density

$$\rho_1 < 2 \times 10^3 (S/10^8)^{-4/3} \text{ g cm}^{-3} \quad (11)$$

But a high-entropy plasma with $S \gg 1$ is known⁶ to be stable against any instability of the type required up to $\rho \approx 10^{14} \text{ g cm}^{-3}$, so there is never any opportunity for fluctuations to form however matter behaves¹⁵ at higher densities. The conclusion is that in a hot universe the macroscopic structure of the universe cannot be explained by any conventional physics—it must originate at an epoch when either classical causality or mass-energy conservation breaks down, and therefore lies in the domain of some theory of 'quantum gravity'. This conclusion (which has been assumed by most authors) remains valid even for 'tepid' universes¹⁶ with a primordial S as low as 10.

Let us consider a cold degenerate universe^{6,7,10,17} with $S \leq 1$. Particle interactions which might cause phase instabilities can be important even at relatively modest densities $\sim 1 \text{ g cm}^{-3}$, as in the models of Zel'dovich and Layzer where the cosmic medium shatters due to solid-body or Van der Waals forces. The maximum redshift for forming fluctuations in a pressureless ($p \ll \rho c^2$) universe is

$$z_{\max} = (M_*/M_H(t_0))^{-14/9} \approx 10^{14} \quad (12)$$

corresponding to a density of

$$\rho_1 \approx 2 \times 10^{13} \text{ g cm}^{-3} \quad (13)$$

so the limits imposed purely by causality present no problem for

this type of model—there is a wide range of z in which fluctuations could be generated. Equation (13) shows that for any primordial S and any assumptions about supernuclear matter, a phase transition at strongly supernuclear density cannot by itself explain cosmic inhomogeneity unless correlated domains form on scales much larger than the horizon—again, requiring either special (non-uniform) initial conditions or a modification of general relativity^{3,28}. We also know, from equation (11), that if a phase transition at nuclear density generates too much entropy (that is $S \geq 10$) it cannot be responsible for cosmic structure, so our model is constrained to remain cool.

One can place more stringent limits on t_1 by considering the process in more detail. In a shattering- or cracking-type phase transition, the size of the lumps created, and hence the distance over which matter is moved, is determined not by the velocity of light, but by the velocity of sound. Thus in equation (4) we should replace M_H by M_S :

$$\delta_M(t_1) \leq (M/M_S)^{-7/6} \quad (14)$$

where

$$(M_S/M_H)_n = (c_s/c)^3 = (p/\rho)^{3/2} \quad (15)$$

Because pressure forces balance cohesive forces in the solid phase, this condition is equivalent to Layzer's: the characteristic size arises because if the expansion energy of a 'fragment' exceeds its cohesion energy, it shatters into still smaller fragments until the two are about equal. Let us assume that the cosmic medium consists of e^- , p^+ , and ν in equal abundance¹⁰, so that the pressure is due mainly to degenerate relativistic electrons and neutrinos. Now $p \approx \rho/3$ when the Fermi energy of the electrons $E_f \geq m_p c^2$, thereafter, p/ρ decreases as a^{-1} until $E_f \sim m_e c^2$, and then like a^{-2} . Thus

$$(M_S/M_H)_n \leq (3a_1/a_2)^{-3/2} \quad (16)$$

where a_2 is the scale factor when $E_f = m_p c^2$ (ref. 14):

$$z_2 = (a_0/a_2) = m_p c n_0^{-1/3} / \hbar (3\pi^2)^{1/3} \approx 8 \times 10^{14} \quad (17)$$

But then equations (14), (7), and (1) say that the sound speed decreases so fast with z that the final fluctuations are maximized if $z \sim z_{\max}$; that is, instabilities at lower or higher redshifts are always fundamentally insufficient to produce the observed clustering, but if a cold universe is unstable at nuclear density the amplitude is about right. (In fact, Zel'dovich and Layzer find that the amplitude of fluctuations from planet-sized lumps resulting from shattering at $\rho \sim 1$ is insufficient and needs to be supplemented with subsequent growth far in excess of the linear growth rate (equation (1)).

Thus, if a phase transition occurs in cold matter at $\rho \sim 10^{13-14} \text{ g cm}^{-3}$ it would yield lumps of order the horizon size at that epoch, the Chandrasekhar mass:

$$M_H = (\hbar c / G m_p^2)^{3/2} m_p \approx 2 M_\odot$$

and would produce $(\delta\rho/\rho) \approx 1$ on present day mass-scales of the order of $10^{12}-10^{15} M_\odot$, possibly enough to explain the observed astronomical inhomogeneity. The surprising thing is that such a phase transition probably exists^{18,31}.

Recent theoretical work on the composition of cold matter in neutron stars indicates that the favoured phase at $\rho \approx 10^{15} \text{ g cm}^{-3}$ is probably a solid Bose lattice of pions held in place by strong nuclear forces¹⁸. At lower densities the Fermi energy of pion decay products decreases, and at some point the equilibrium state becomes a Fermi gas of baryons and leptons. An additional uncertainty is introduced in cosmological cold matter because the relative chemical potentials of fermion species depend on the cosmological lepton-baryon ratio n_L/n_B (in neutron stars lepton number is not conserved because neutrinos can escape). Although more calculations are necessary to include this effect, it seems likely that a phase transition from pion condensate to Fermi gas occurs for at least some values of n_L/n_B . This phase transition is of the right type, and occurs at about the unique right time as calculated above, to produce the observed astronomical inhomogeneity.

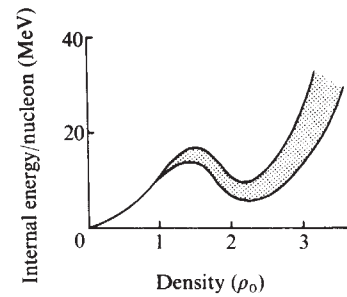


Fig. 1 Equation of state of cold matter around nuclear density.

Figure 1, after Irvine¹⁸, shows the specific internal energy of cold matter $E(\rho)$ near nuclear density $\rho_0 \approx 3 \times 10^{14} \text{ g cm}^{-3}$. There are theoretical uncertainties (shaded region), but it is clear that at zero temperature the conditions for thermodynamic instability²⁹ exist:

$$\begin{aligned} (\partial E / \partial \rho) &< 0 \\ (\partial^2 E / \partial \rho^2) &< 0 \end{aligned} \quad (18)$$

which require the solid Bose condensate to shatter at $\sim 2\rho_0$ into macroscopic fragments. (Note that matter with $S \geq 1$ would not have this property: the internal energy then varies monotonically with ρ because it is dominated by thermal photons and neutrinos.) Chaotic conditions and large-scale motions persist for about a Hubble time, leaving lumpy single-phase matter with $\delta\rho/\rho \geq 1$ at a mean cosmic density of the order of $\rho_1 \sim 10^{13-14} \text{ g cm}^{-3}$.

As the Jeans mass in a cold universe decreases with time after t_2 , it is probable that fragments would collapse and survive as compact bound objects, providing a natural source of 'missing mass'. We also know that the fragments are initially of the order of a Chandrasekhar mass—they might end up as black holes, neutron stars, main-sequence hydrogen-burning stars or super-massive stars ($\geq 100 M_\odot$), depending on the exact details of the fragmentation and subsequent coalescence. Such objects are natural candidates for the energy sources necessary to explain the microwave background in a cold model¹⁹⁻²¹.

There are some weak points in the above hypothesis which also provide crucial observational tests. Some potential problems are: (1) a fluctuation spectrum as steep as equation (2) cannot explain both galaxies and galaxy clusters if only gravitational forces are at work⁸, it forms galaxies too early and hence with too high a density; (2) it is not certain that enough inhomogeneity on galaxy clustering scales $\approx 10^{15} M_\odot$ can be generated with only the linear growth rate (equation (1)); and (3) the cosmic helium abundance cannot be calculated—even if n_L/n_B were known, the chaotic conditions caused by the phase transition would invalidate previous cold-universe calculations^{10,22} by increasing the entropy and upsetting the equilibrium abundance of various particle species (for example, collapsing fragments would squeeze neutrinos into the confining Fermi gas). The first two of these problems can be solved by invoking hydrodynamic effects, for example, primordial fragments may form into super-massive exploding stars^{21,30} which could unbind dense systems on scales $\leq 10^{12} M_\odot$ and/or amplify fluctuations on scales $10^{12} \leq M \leq 10^{15} M_\odot$. Such objects would certainly leave an imprint on the anisotropy²³ and spectrum²⁴⁻²⁶ of the microwave background; they might also create helium and/or deuterium²⁷. A fairly rigorous prediction of any shattering model is that present-day fluctuations on scales $\gg 10 \text{ Mpc}$ should be unaffected by such events and should still show the original $(\delta\rho/\rho) \propto M^{-7/6}$ spectrum.

Although there are complications with a cold big bang, it is clear from the preceding discussion that any other model attempting to produce structure in an initially uniform universe is likely to encounter difficulties of a much more fundamental nature unless it tampers with classical general relativity. If the

structure we see today is not the shattered remnant of a cold universe, then it must be a relic of the preclassical era.

I thank Bernard Carr, David Layzer, P. James E. Peebles and Martin Rees for helpful comments, and the University of Cambridge for an Isaac Newton Studentship.

Received 25 February; accepted 3 June 1980.

- Harrison, E. R. *Phys. Rev. D*, **1**, 2726 (1970); *Nature* **215**, 151 (1967).
- Peebles, P. J. E. *Nature* **220**, 227 (1968).
- Press, W. H. *Phys. Scr.* **21** (1980).
- Sato, K. Preprint (Nordita, Copenhagen, 1980).
- Lasher, G. *Phys. Rev. Lett.* **42**, 1646 (1979).
- Layzer, D. in *Galaxies and the Universe* (eds Sandage, A., Sandage, M. & Kristian, J.) (University of Chicago Press, 1976).
- Zel'dovich, Ya. B. *Soviet Phys. JETP* **16**, 1395 (1963); *JETP* **43**, 1982 (1962).
- Peebles, P. J. E. *Cosmology: Physics of Large-Scale Structure* (Princeton University Press, 1980).
- Hawking, S. W. & Ellis, G. F. R. *The Large Scale Structure of Space-Time* (Cambridge University Press, 1973).
- Zel'dovich, Ya. B. *Adv. Astr. Astrophys.* **3**, 241 (1965).
- Peebles, P. J. E. *Astr. Astrophys.* **32**, 391 (1974).
- Press, W. H. & Vishniac, E. T. *Astrophys. J.* (submitted).
- Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
- Zel'dovich, Ya. B. & Novikov, I. D. *Relativistic Astrophysics*, Vol. 1 (eds Thorne, K. S. & Arnett, W. D.) (Midway, University of Chicago Press, 1978).
- Weinberg, S. *Phys. Scr.* **21**, 773 (1980).
- Carr, B. J. & Rees, M. J. *Astr. Astrophys.* **16**, 705 (1977).
- Carr, B. J. *Mon. Not. R. astr. Soc.* **181**, 293 (1977).
- Irvine, M. *Neutron Stars* (Clarendon, Oxford, 1978).
- Layzer, D. & Hively, R. M. *Astrophys. J.* **179**, 361 (1973).
- Rees, M. J. *Nature* **275**, 35 (1978).
- Carr, B. J. Preprint (University of Cambridge, 1980).
- Kaufman, M. *Astrophys. J.* **160**, 459 (1970).
- Hogan, C. *Mon. Not. R. astr. Soc.* (in the press).
- Woody, D. P. & Richards, P. L. *Phys. Rev. Lett.* **42**, 925 (1979).
- Rowan-Robinson, M., Negroponte, J. & Silk, J. *Nature* **281**, 635 (1979).
- Puget, J. L. & Heyvaerts, J. *Astr. Astrophys.* **83**, L10 (1980).
- Epstein, R. I. *Astrophys. J.* **212**, 595 (1977).
- Zee, A. *Phys. Rev. Lett.* **44**, 703 (1980).
- Landau, L. D. & Lifshitz, E. M. *Statistical Physics* (Pergamon, Oxford, 1969).
- Doroshkevich, A. G., Zel'dovich, Ya. B. & Novikov, I. D. *Soviet Astr.* **11**, 233 (1967); *Astr. Zh.* **44**, 295 (1967).
- Shuryak, E. V. *Phys. Rep.* **61**, 71–158 (1980).
- Bardeen, J. M. Preprint (Institute for Advanced Study, Princeton, 1980).

Some diurnal properties of clouds over the martian volcanoes

Garry E. Hunt*, Adrian O. Pickersgill*,
Philip B. James† & Gordon Johnson‡

*Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

†Department of Physics, University of Missouri, St Louis, Missouri 63121

‡Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91103

Discrete white cloud systems have long been recognized as a martian atmospheric phenomenon. However, it was not until the Mariner 9 mission in 1971–72 that the cloud systems were found to be closely related to topographic features, in particular, the giant volcanoes on the Tharsis Ridge. Using the available Earth based observations, Smith and Smith¹ investigated the seasonal and diurnal behaviour of the clouds the Olympus Mons and Elysium regions. They found that the cloud activity was generally confined to late spring and early summer seasons in the northern hemisphere. This corresponds to $L_s = 60$ – 150° , where L_s is the aerocentric solar longitude and $L_s = 0$ for spring equinox in the northern hemisphere. The Viking Orbiters have observed martian atmospheric phenomena for eight seasons. During this period, the volcanoes have been frequently, but often irregularly, monitored. Unfortunately, the changes in the spacecraft orbit have meant corresponding variations in the surface resolution, viewing geometry and local times of the observations. However, the available observations do provide important new insights into the diurnal properties of the volcanic cloud systems, which we describe here.

Cloud systems appear over the volcanoes in certain seasons of the martian year, and these spacecraft observations are in general agreement with the times of meteorological activity suggested by the ground based studies of Smith and Smith¹. In the Elysium region, clouds were observed during the period $L_s = 46$ – 141° , while during $L_s = 230$ – 17° there was no apparent atmospheric activity. Except for a period between $L_s = 2100$ and $L_s = 300^\circ$, some form of cloud activity was observed throughout the year in the Tharsis region. However, this period of apparent inactivity corresponds to the time of two major dust storm events observed by the Viking spacecraft (see, for example, ref. 2). This area of the planet was noticeably dusty, and frequently only the summits of these volcanoes were visible through the dust clouds. During these dust storms, the atmospheric water vapour abundance was greatly reduced³ as the particles absorbed the atmospheric moisture. Furthermore, the thermal properties of the atmosphere were also affected by the presence of the dust, suppressing the diurnal variation and reducing the vertical structure^{4,5}. These substantial changes to the thermodynamic state of the atmosphere account for the absence of clouds due to the dust storm period.

The cloud systems observed from the Earth cover areas of 10^5 – 10^6 km² and are first seen at periods around local noon. From Viking studies, we have found many much smaller clouds in the Tharsis region, occupying regions of only $\sim 5 \times 10^3$ km², which often develop in the early morning. Figure 1 shows such a localized cloud system for the Ascreaus region at $L_s = 225^\circ$. Here we concentrate on a study of the diurnal properties of orographic cloud systems, monitored near Elysium, Olympus Mons, and the two northern peaks of the Tharsis system.

Figure 2 shows the development of a cloud system associated with Elysium from 07.00 to 12.51 LT at $L_s = 92.7^\circ$, shortly after summer solstice. All of the clouds seen here are smaller than the classic afternoon clouds seen from the Earth. Each of the views in Fig. 2 were imaged in four filters. By comparing these images condensate clouds can be separated from surface details produced by slopes which appear as bright areas through all filters. In the first three pictures, the sunward facing slopes are brightly highlighted in a manner frequently observed at Elysium. The clouded western flanks show greatly reduced contrast as the wavelength is increased, unlike the sunward slopes which appear equally bright at all wavelengths. At 07.00 LT, the entire region is lightly clouded and surface features may be seen



Fig. 1 Viking Orbiter 1, revolution 225, frame 5. A turbulent region of cloud is observed on the western flanks of Ascreaus Mons for 14.03 LT through a violet filter for $L_s = 210^\circ$.

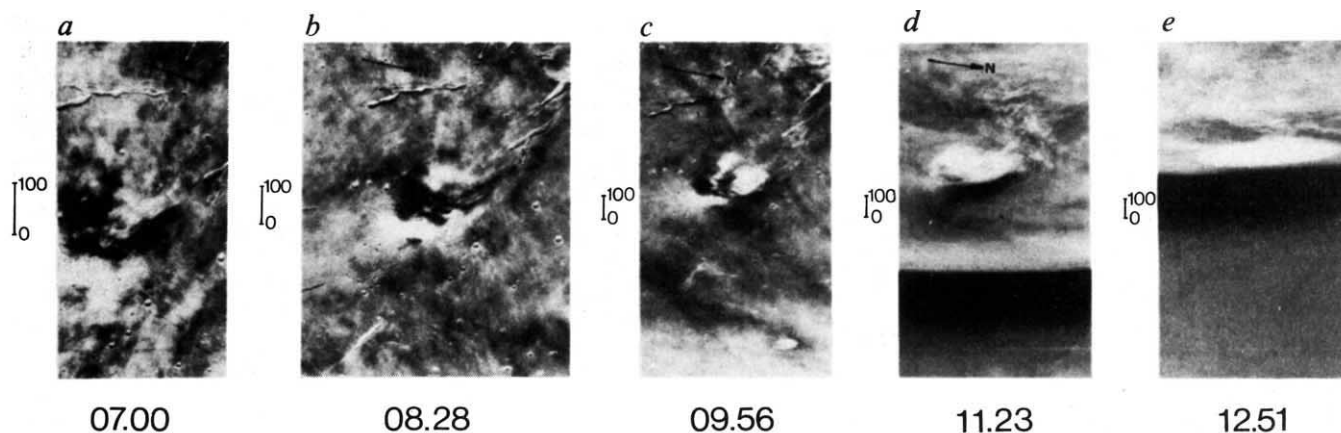


Fig. 2 The daily cycle of a cloud in the lee of Elysium Mons is shown as five time-sequence photographs taken between 07.00 and 12.51 LT. The LT of the frames are 07.00 (a), 08.28 (b), 09.56 (c), 11.23 (d) and 12.51 (e). The arrow-shaped lee wave cloud on the extreme left photograph is replaced by an almost circular region of cloud on the extreme right photograph. These photographs were taken by Viking Orbiter 1, orbit 710, through a violet filter for $L_s = 92.7^\circ$.

through the denser patches, although with reduced contrast. The most prominent cloud is seen to the WNW of Elysium in the vague shape of an arrow whose head points towards the summit of the volcano. This cloud is more opaque than the mist in the surrounding region. The region has apparently not changed appreciably in the second view, obtained at 08.28 LT. The cloud to the WNW is still quite apparent although it is more concentrated than in the earlier view. By 09.56 LT (Fig. 2c), this cloud has clearly split into two pieces, with the main portion displaced substantially towards the north, while a plume, reminiscent of those seen in the Tharsis region⁶ appears to the west. The larger component of the cloud is roughly 100 km in diameter.

The cloud's rapid growth seems to be starting in the 11.23 view shown in Fig. 2d. The dimensions of this cloud are roughly 200 km, so that the effective area has roughly quadrupled in the hour and a half since the previous picture. The region is somewhat hazy, and a field of scattered clouds seems to be developing to the north-west of the main cloud. The approximately 70° emission angles roughly double the effective optical depth from the earlier view and, therefore, enhance the haziness of the region. In the final view 12.51 LT, the cloud has grown to dimensions of ~ 350 km.

Figure 3 summarizes a diurnal sequence in the vicinity of Ascræus Mons for $L_s = 121.5^\circ$. The first images, corresponding to 05.50 and 07.30 LT show a pervasive fog which nearly obscures surface features on the flanks of Ascræus. Comparison with the USGS topographic map of Mars suggests that the top of the cloud bank is ~ 15 km below the summit so that this is a very low lying cloud layer.

The cirrus-like clouds to the south-west of the summit in Fig. 3a seem to be of the variety catalogued by Briggs *et al.*⁶ one (martian) year earlier. They are well over the terminator and, on the basis of their illumination, would have to be at least as high as the summit, which is 27 km. Two discrete clouds are seen to the west of the volcano in Fig. 3b. Ascræus frequently exhibits clouds in this location and they sometimes take the form of extensive plumes such as those reported by Briggs *et al.*⁶. Apparent shadows suggest that these may be quite high, perhaps 10–15 km above the volcano summit. This feature is confined to the early morning hours at all seasons, and is a lee wave cloud system formed on the leeward side of the volcano. These features imply the existence of easterly winds at high altitudes in agreement with the predictions of circulation models^{7,8}.

The general cloudiness at low levels is still seen in Fig. 3c, obtained when the local time is 09.26 at the summit of Ascræus. However, a variety of waves are present with wavelengths around 10 km. The orientation of the waves suggests streamlined flow around the volcano, but the presence of waves at opposed angles may imply a more complicated situation. There

are no obvious surface features that could affect the flow and be identified as the source of the waves. However, in view of the wide variety of undulations on the planet, it is still possible that these waves are topographically induced. The time localization of these phenomena suggests that they may be the result of a change in the characteristics of the boundary layer which is strongly influenced by the thermal properties and opacity of the atmosphere. Bright areas to the east of the volcanoes are probably incidence angle effects since they have disappeared in the later view. It is not possible to make a more positive statement because the images are available only in the violet filter (450 nm).

By 11.23 LT (Fig. 3d) large clouds are seen developing to the west and south of Ascræus and to the west of Pavonis, which suggests that these are the nuclei of two components of the w cloud monitored by the Earth based observers (see, for example, ref. 11). Cloudiness is also seen to the east of Ascræus towards the canyons and the clouds in these peripheral areas appear distinctively convective. Estimates of heights from other pictures using presumed shadows suggest that these clouds are relatively low with altitudes of 4.5–6.5 km above the surface.

The present visible observations are supported from the Orbiter measurements of the thermal characteristics and water vapour distribution of the atmosphere. This then allows the composition of the clouds to be determined by the procedures developed by G.E.H.⁹. We find that the low lying fogs and the convective afternoon clouds (see also ref. 10) are water ice clouds. In this season, the atmospheric temperatures are far above the CO_2 condensation temperatures. The cirrus-type clouds are more difficult to characterize owing to their very diffuse nature. Certainly, as Hunt¹¹ has shown, we cannot yet exclude the possibility of CO_2 clouds at these high altitudes.

The diurnal development of clouds, shown in Figs 2 and 3, is related to the local stability and circulation of the atmosphere as influenced by the topography and local thermal properties of the surface. The volcanic areas have low surface inertias and this permits unusually large diurnal temperature variations in these areas¹⁰. Night-time cooling produces denser cold air at low elevations favouring the formation of the observed radiation fogs in the early hours of the morning. The water vapour content of the martian atmosphere seems to be limited by the amount which can be contained in a saturated, night-time atmosphere¹². Regions of locally large cooling are, therefore, favoured sights for fogs. The atmospheric stability associated with these conditions, together with the downslope winds, favour early morning gravitational wave effects such as are seen particularly in the Tharsis region.

The rapid warming of the surface as noon approaches will lead to increased lapse rates and, perhaps, to eventual instability.

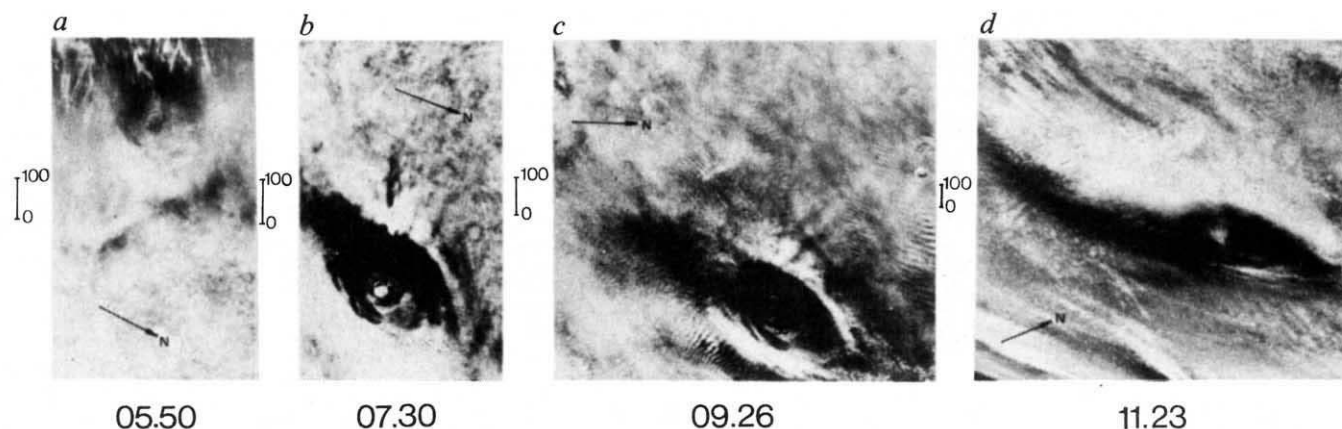


Fig. 3 The morning cycle of cloud in the vicinity of Ascraeus Mons shown as four time-sequenced photographs, taken between 05.50 and 11.23. The LT of the frames are 05.50 (a), 07.30 (b), 09.26 (c) and 11.23 (d). The scale and direction of north are shown. These photographs were taken by Viking Orbiter 1, orbit 774, through violet filter for $L_s = 121.5^\circ$.

Lindal *et al.*¹⁴ have noted that lapse rates approach adiabatic only in these latitudes and seasons. The resulting convection will produce cumulus-like clouds by adiabatic cooling; this is supported by observations of such clouds in the region. Upward circulation in the atmosphere has been predicted to the west of the large volcanoes due to thermal forcing effects¹³. Cloud formation will also be enhanced by this convergence due to the positive vertical velocities; winds from the west due to this low level convergence may also contribute to adiabatic cooling of the parcels of air as they are carried up the slope of the Tharsis Ridge. The combination of these three effects seems plausible as an explanation for the diurnal cycle observed.

The reason for the pronounced seasonal variation is less easily seen, however. The spring and summer seasons in the northern hemisphere correspond generally to higher insolation and larger water vapour content in the atmosphere than does the rest of the year. As both factors should influence cloud formation, the general correlation is not surprising. However, in detail these factors fail to explain the abrupt boundaries of activity at $L_s = 60^\circ$ and $L_s = 150^\circ$ for Ascraeus and Elysium. The water at 25°N peaks at L_s between 140° and 150° , when the cloud activity is ending, and only gradually drops to equinox¹⁵. The large martian orbital eccentricity causes the insolation at noon at 25° latitude to be actually greater at the equinoxes than at solstice. Possibly, circulation changes when the south cap becomes stationary at $L_s \sim 150^\circ$ may effect the change; such circulation changes, from easterly to westerly winds, are predicted at that time by GCM⁷. If this change from easterly to westerly winds occurs the formation of convective clouds to the west of the volcanoes may be inhibited.

The Tharsis volcanoes are cloudy for a more extensive period and this is consistent with all these factors as the circulation and insolation at the equator do not vary as they do to the north. The cessation of activity from $L_s = 210^\circ$ to $L_s = 300^\circ$ is clearly associated with the major dust storms. These not only reduce the water content of the atmosphere drastically, but also tend to make the vertical profiles more isothermal than would normally be expected.

G.E.H. and A.O.P. were supported by the SRC. P.B.J. and G.J. were supported by NASA.

8. Pollack, J. B., Leovy, C. B., Mintz, Y. H. & Van Camp, W. *Geophys. Res. Lett.* **3**, 479–482 (1979).
9. Hunt, G. E. *J. geophys. Res.* **84**, 2865–2874 (1979).
10. Kieffer, H. H. *et al. J. geophys. Res.* **82**, 4249–4291 (1977).
11. Hunt, G. E. *Geophys. Res. Lett.* (in the press).
12. Davis, D. W. *J. geophys. Res.* **84**, 2875–2879 (1979).
13. Webster, P. J. *Icarus* **30**, 626–649 (1977).
14. Lindal, G. *et al. J. geophys. Res.* **84**, 8442–8457 (1979).
15. Farmer, C. B., Davies, D. W., Holland, A. L., La Porte, D. D. & Doms, P. E. *J. geophys. Res.* **82**, 4225–4248 (1977).

Black-body radiation as an inertial confinement fusion driver

F. Winterberg

Desert Research Institute, University of Nevada System, Reno, Nevada 89507

To ignite a thermonuclear microexplosion a power source of $\geq 10^{14}\text{W}$ is required with an energy of $\geq 10^6\text{J}$ focused down to $\leq 0.1\text{cm}$. To meet these driver requirements, charged particle beams¹ and even hypervelocity projectiles² have been proposed as an alternative to lasers for which an energy $\geq 10^6\text{J}$ is considered very large. Although charged particle beams and hypervelocity projectiles promise much larger energies the problem of the large driver power of $\geq 10^{14}\text{W}$ remains. We propose here a two-stage driver concept which eliminates this problem. In it the initial energy is drawn from any of these drivers but is used here to generate an intense black-body radiation source. This black-body radiation is then used to drive the thermonuclear microexplosion target. This concept not only permits the initial driver power to be reduced by about two orders of magnitude from $\geq 10^{14}\text{W}$ down to a few 10^{12}W , but also greatly relaxes the beam-focusing requirements as compared with direct pellet fusion.

The idea is depicted in Fig. 1. The thermonuclear target T is placed inside the centre of a small cavity filled with a tenuous high atomic number gas, for example, krypton or uranium hexafluoride. The inside of the wall is lined with a dense material P, for example gold or uranium. P is surrounded by an ablative layer A of low atomic number material. Beams B hitting the ablator from all sides ablatively implode the cavity. The implosion makes the wall P implode with the velocity v into the tenuous gas inside the cavity. This results in a shock

Received 30 April; accepted 26 June 1980.

1. Smith, S. A. & Smith, B. A. *Icarus* **16**, 509–521 (1972).
2. Briggs, G. A., Baum, W. A. & Barnes, J. *J. geophys. Res.* **84**, 2795–2820 (1979).
3. Farmer, C. B. & Doms, P. E. *J. geophys. Res.* **84**, 2881–2898 (1979).
4. Martin, T. Z., Peterfreund, A. R., Miner, E. D., Kieffer, H. H. & Hunt, G. E. *J. geophys. Res.* **84**, 2830–2842 (1979).
5. Martin, T. Z. & Kieffer, H. H. *J. geophys. Res.* **84**, 2843–2852 (1979).
6. Briggs, G. A., Klaassen, K., Thorpe, T., Wellman, J. & Baum, W. *J. geophys. Res.* **82**, 4121–4150 (1977).
7. Leovy, C. B. & Mintz, Y. *J. Atmos. Sci.* **26**, 1167–1190 (1969).

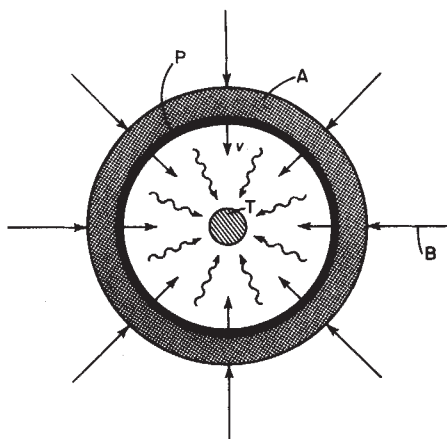


Fig. 1 The two-stage driver equipment: T, thermonuclear target; P, lining of dense material surrounded by an ablative layer A. Beams B, hit the ablator from all sides.

wave moving into this gas. The temperature T behind this shock wave is related to the implosion velocity³ and one finds

$$T = \frac{A}{3R(Z+1)} v^2 \quad (1)$$

where A is the atomic weight, R the gas constant and Z the degree of ionization. Z depends on T and a useful approximation is

$$Z \simeq 6.6 \times 10^{-3} T^{0.41} \quad (2)$$

The shock-heated high temperature gas, by bound-free, bound-bound and free-bound transitions, will then become an intense source of photons, and as a result the kinetic energy in the shock wave will be rapidly converted into black-body radiation. This happens on an atomic time scale of $\lesssim 10^{-8}$ s. The resulting black-body radiation is further compressed and greatly amplified by the cavity implosion, and is finally delivered to the target surface, assumed to be spherical of radius r_p at the rate

$$P(t) = 4\pi r_p^2 \sigma T^4 \quad (3)$$

where σ is the Stefan-Boltzmann constant. A power of 10^{14} W for a pellet surface of $\sim 10^{-1}$ cm² is reached at a temperature of $\simeq 3.6 \times 10^6$ K. To store energy of $\simeq 10^7$ J at this temperature in the form of black-body radiation would require a volume of 75 cm³; this suggests we use higher temperatures, as for a given energy the volume decreases as T^{-4} . For $T = 10^7$ K the cavity volume would be only $\simeq 1.3$ cm³. In the absence of photon losses through the cavity wall the black-body radiation changes with the cavity volume V as $VT^3 = \text{constant}$, or with the cavity radius r as $rT = \text{constant}$. Therefore, if the initial cavity temperature is $T = 5 \times 10^6$ K, a twofold reduction in the cavity radius would raise the temperature to 10^7 K with a 2^4 fold increase in the radiative power delivered to the thermonuclear target. The initial and final radiative power would thus be $\simeq 3.6 \times 10^{14}$ W and 5.75×10^{15} W. From equations (1) and (2) to reach the initial temperature would require $A=200$, $v=46$ km s⁻¹ and krypton with $A=83$, $v \simeq 71$ km s⁻¹.

To make a shock wave in the tenuous gas the collision mean free path must be small compared with the cavity radius and one finds that the atomic number density must be $\gtrsim 10^{18}$ cm⁻³. On the other hand, to store most of the incident energy into black-body radiation, the gas density cannot exceed $\sim 10^{21}$ cm⁻³. Finally, if the thermonuclear target were to be bombarded with the full power derived from the Stefan-Boltzmann law, the cavity would have to be optically

transparent, because otherwise the photons would diffuse through the gas at a much lower average velocity. Whether or not the gas is optically transparent is determined by the opacity coefficient⁴

$$\kappa \simeq 4.34 \times 10^{25} \rho T^{-3.5} \quad (4)$$

from which the photon path length $\lambda_p = (\kappa \rho)^{-1}$ is derived. To make $\lambda_p > r \sim 1$ cm requires $\rho \lesssim 0.3$ g cm⁻³, which for $A=200$ is an atomic number density of $\lesssim 10^{21}$ cm⁻³. Therefore, an atomic number density of $\simeq 10^{19}$ cm⁻³ would satisfy all the conditions.

The concept depends on the black-body radiation photons being sufficiently confined during the cavity implosion. The diffusion of the photons through the cavity wall is given by the energy flux

$$j = \frac{\lambda_p c}{3} \frac{\partial}{\partial x} (aT^4) \quad (5)$$

where $a \equiv 4\sigma/c$, λ_p is the photon path length in the wall, which like the path length in the cavity must be computed using equation (4). Putting $j = aT^4 v_d$ and $\partial(aT^4/\partial x) \simeq aT^4/x$ where x is the distance the photons have travelled into the wall with the diffusion velocity v_d , one finds

$$v_d/c \simeq (1/3)(\lambda_p/x) \quad (6)$$

For $T = 10^7$ K, $\rho = 18$ g cm⁻³ (uranium) one finds $\lambda_p \simeq 2 \times 10^{-4}$ cm, and requiring $x \lesssim r \sim 1$ cm, that $v_d > 20$ km s⁻¹. Therefore, a cavity implosion velocity of $v \gtrsim 50$ km s⁻¹ should be sufficient to confine the photons effectively.

Because the radiation flux absorbed by the thermonuclear target is inversely proportional to the opacity of the target and which according to equation (4) is proportional to $T^{-3.5}$, the power absorbed will rise in proportion to $T^{7.5}$ or $r^{-7.5}$. If, for example, the cavity radius decreases twofold, the power absorbed increases by the factor $2^{7.5} \simeq 180$. Most of the energy is, therefore, delivered to the pellet at peak power towards the end of the implosion process. This power amplification is due to the initial driver energy intermediately stored here in the form of black-body radiation which is only delivered at the end of the implosion process. Because of this the initial driver power can be reduced by a factor of about 100.

The short wavelength of the black-body radiation, given by $h\nu = kT$, and which is in the soft X-ray domain, is much better

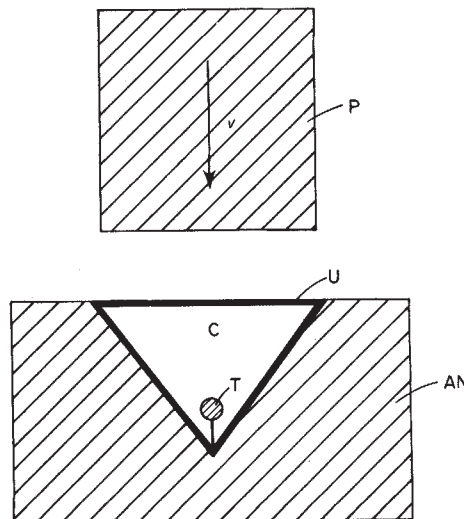


Fig. 2 A combination of the two-driver process with impact fusion. P, a hypervelocity projectile; C, conical cavity carved out of the anvil, AN; T, thermonuclear target.

suited than laser light to reach high density target compression, as it is matched to a plasma frequency of $\sim 10^4$ times compressed hydrogen. Furthermore, because of the isotropy of the radiation, there is no stimulated Brillouin backscattering as in laser fusion. A likely candidate to drive the cavity implosion seems to be an intense light ion beam (which can already be produced at the required power level) which is sufficiently well focused because the target area of the cavity is much larger than that of the pellet.

The rather low implosion velocity of $\sim 50 \text{ km s}^{-1}$ suggests combining this concept with impact fusion where a hypervelocity projectile P accelerated to $v \sim 50 \text{ km s}^{-1}$ implodes and amplifies black-body radiation in a conical cavity C carved out the anvil AN with the thermonuclear target T inside this same cavity. This concept is shown in Fig. 2.

Received 5 March; accepted 23 April 1980.

1. Yonas, G. *Scient. Am.* **239**, 50 (1978).
2. *Proc. Impact Fusion Workshop*, Los Alamos (1979).
3. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics*, 331, 358 (Pergamon, London, 1959).
4. Schwarzschild, M. *Structure and Evolution of Stars*, 67 ff (Princeton University Press, 1958).

Overlap of intense charged particle beams for inertial confinement fusion

J. A. Halbleib, P. A. Miller, L. P. Mix & T. P. Wright

Sandia National Laboratories, Albuquerque, New Mexico 87185

One approach to controlled thermonuclear fusion for power production requires the irradiation of centimetre-size targets containing heavy hydrogen (D-T) fuel by intense charged particle beams (CPB). Beam energy and pulse length requirements are $\sim 1 \text{ MJ}$ and 10 ns . In certain studies¹ the CPBs are beams of light ions, such as helium, with kinetic energies per nucleon in the range $1\text{--}5 \text{ MeV}$. These beams can be transported by preformed, current-carrying, plasma discharge channels. The fields from the channel currents (supplied by capacitor banks) confine the CPBs. These channels can be laser-initiated in a background gas that has a pressure between 10 and 100 torr. This gas blanket enables first wall loading, by X rays, shocks and thermal radiation to be mitigated. Although intense relativistic electron beams (REBs), which were developed before intense light ion beams, have achieved higher intensities², ion beams have more favourable deposition characteristics in targets³. Proton beams have driven exploding pusher targets⁴ to implosion velocities of $20 \text{ cm } \mu\text{s}^{-1}$ and have attained power of $\sim 1 \text{ TW cm}^{-2}$, which is a factor of 40 below that estimated (G. Allhouse, personal communication) to be necessary for target ignition and for net system energy gain. To close the gap between present capabilities and future requirements, beam bunching and beam overlap are being studied. The current density gain G is important to the CPB overlap process and is the ratio of beam current density at the target surface to that of the individual beams in the transport channels; it depends on particle species, transverse energy, geometry of the channel-target interface, and so on. Detailed calculations⁵⁻⁷ indicate that for electron beams, $G < 3$, while for practical cases of light ions, G can exceed 10. This is due to the difference in masses and effective temperatures of the two types of beams. Sources providing multiple ion beams are not available to test this overlap theory experimentally, hence it has been tested using a multiple electron beam system. Here we outline the experiment and describe the apparatus, and then summarize the theoretical treatment of beam overlap, describe diagnostic procedures, and compare the experimental results with calculations.

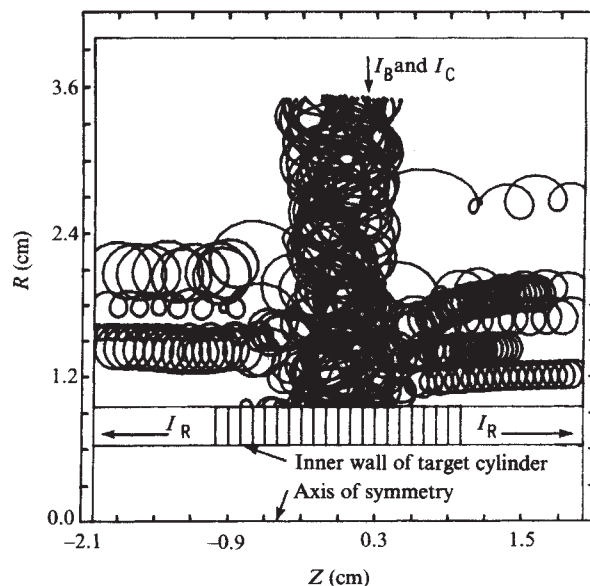


Fig. 1 R - Z trajectory plot for beam electrons incident on a 1.9 cm o.d. target cylinder. Target zoning for deposition analysis is indicated but trajectories inside the target are suppressed. I_B , I_C , and I_R are magnitudes of the beam currents, channel conduction currents and return currents ($3I_C$), respectively; directions of the electron flows are indicated by the arrows.

Several electron beams were generated simultaneously and transported through plasma channels^{8,9} to a common cylindrical target. Targets with different diameters were used, and the total energy delivered to the targets was measured. As the diameter of the target decreases, progressively less of the energy reaches the target. The experiment aimed to measure the dependence of delivered energy on target diameter and to compare this with theoretical predictions. Agreement would imply that the physical models are correct in the present energy density regime, however, further work would still be required to investigate collective effects which may be important with higher-energy density systems^{1,10-12}.

The electron beam source was the Proto II accelerator¹³ at Sandia National Laboratories. The front end of the accelerator could drive 12 separate electron beam diodes in parallel. These diodes were arranged in coplanar fashion in a 1-m diameter circle. On each accelerator shot, 12 pinched electron beams were generated, with six of these being stopped by 0.64 cm-thick solid aluminium or copper anodes. The other six were extracted from the diodes through thin compound diaphragms in the centres of the anodes. The six extracted beams were transported by preformed plasma channels of which the target was the common terminus. The channels were formed in parallel by the discharge of a $20 \mu\text{F}$ capacitor bank system charged to 65 kV. System inductance was 330 nH. The six plasma channels used for beam transport were formed in the ambient laboratory atmosphere and were guided by $25 \mu\text{m}$ diameter tungsten wires.

The phenomenology and limitations of plasma channel transport have been discussed in more detail elsewhere¹. Basically, the channels start with initial diameters $< 1 \text{ mm}$ and expand with radial velocities of $\sim 1.5 \text{ mm } \mu\text{s}^{-1}$. The central part of each channel consists of singly ionized air at a few electron volts temperature, with a mass density reduced below ambient by a factor near four. There is much evidence that the tungsten wire plays a role only in initiating the discharge, and that it does not 'explode' nor affect subsequent channel dynamics. Thus the channels' features should be the same as if they had been laser-initiated. As the channels expand, the current rises to a peak in $4 \mu\text{s}$; at this time the accelerator is fired. The conductivity of the channels is high enough to provide space-charge and current neutralization of the beam fields. Thus, ideally there are

no collective effects, and the beam is guided only by the channel fields from the channel current. Consequently the peak current in each channel must be the Alfvén current, which is ~ 65 kA for the 1.5 MeV electrons.

For this experiment, Proto II was operated in a mode that provides a pulse with nominal parameters of 1.5 MV, 3.5 MA, 20 ns FWHM, 5 TW, and 100 kJ. Thus the energy and power of each individual beam were ~ 8 kJ and 0.4 TW, and were transported 0.46 m. The aluminium targets had diameters ranging from 6 to 32 mm.

The time independent theoretical model for channel transport is based on an earlier technique¹⁴ which combines coupled electron/photon collisional transport with electron transport in static, externally applied electric and magnetic fields of arbitrary spatial dependence. As the relatively short pulse REBs are not injected until the highly conducting plasma channels have been well established, the electrons of the charge and current neutralized beam are assumed to follow single particle trajectories in the quasi-static externally applied field of the channel conduction current. While the material geometry in the numerical model is constrained to be two-dimensional, cylindrical, and invariant with respect to rotation about an axis of symmetry, particle trajectory descriptions are fully three-dimensional. This three-dimensional capability is essential in the present application because of the azimuthal dependence of the multibeam source and because of the preferential orientation of the X-ray pinhole camera whose exposure environment is to be predicted by the model.

Figure 1 shows an R - Z plot of a random sample of trajectories for a 1.91-cm o.d., 1.27-cm i.d. cylindrical target. Only that portion of the electron trajectories outside the target is plotted. Experiments have confirmed efficient transmission along the plasma channels⁶. Consequently, trajectories are initiated in the model calculations at a distance of one betatron wavelength (~ 2.3 cm) beyond the geometrical beam 'overlap radius' (~ 1.2 cm) to ensure spatial phase mixing¹⁵. The overlap radius is defined as the radius at which the edges of the individual beams and channels start to merge. The magnetic field is a superposition of the fields due to the six-channel conduction currents and the axial return currents in the target cylinder. Field effects are ignored inside the solid target because of the high collision frequency. On the other hand, collisional effects are ignored outside the target.

Because the target radius is about three-quarters of the geometrical overlap radius in this calculation, $\sim 100\%$ of the beam strikes the target⁵. The beam/target coupling efficiency (per cent of the beam energy deposited in the target) is about 80%. Fig. 1 shows most of this energy is deposited within the

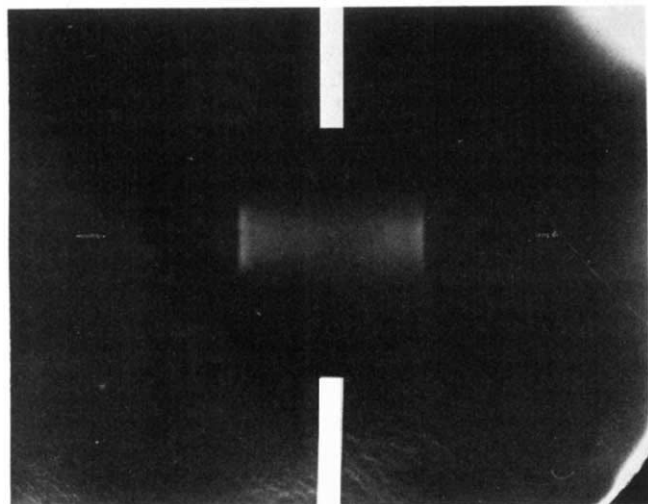


Fig. 2 X-ray pinhole photograph of six beams incident on a 1.9-cm o.d. target. The vertical axis of the cylinder is indicated.

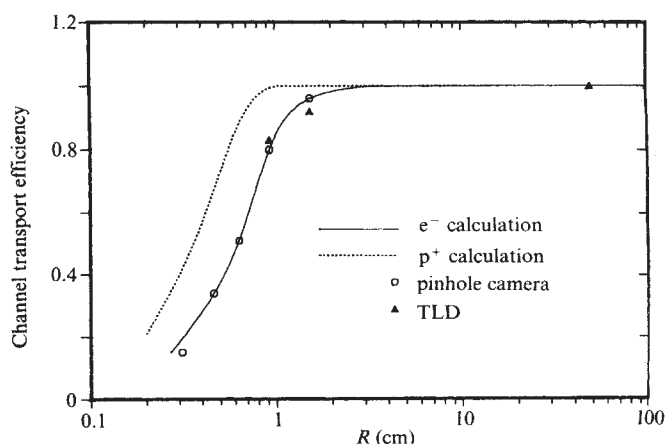


Fig. 3 Beam-target coupling efficiency as a function of target radius. For the 2-MeV proton reference calculation, the maximum angle of injection of protons into the channels was taken to be 8° .

beam radius. A few of the reflected electrons reenter the target. However, the majority of the reflected electrons ($\sim 20\%$ of the beam energy) is lost through axial drift in the field of the return currents.

The numerical model was also used to predict the coupling efficiencies and three-dimensional energy deposition profiles as a function of the target o.d. In addition, spatially and spectrally resolved predictions of the X radiation emitted within a small solid angle about a direction corresponding to the orientation of an experimental pinhole camera were obtained. According to the calculated results, the total X-ray emission in the direction of the pinhole should be proportional to the coupling efficiency.

The fraction of the beam energy transported to the central target cylinder was determined from measurements of the bremsstrahlung produced by the electron beams. An X-ray pinhole camera in the plane of the transport channels viewed the cylinder transverse to its axis. Thermoluminescent dosimeters (TLDs) were positioned at one end of the central target cylinder. Additional TLDs were placed behind the solid aluminium anodes of those beams which were not transported to provide a reference measurement of the dose expected when 100% of the electrons were stopped in aluminium.

Both diagnostics relied on the theoretical calculations presented above for interpretation of the results. The absolute TLD measurements were calculated to depend strongly on the kinetic energy of the electrons striking the aluminium (roughly as the square of the energy). The uncertainties in the absolute voltage monitor calibrations would have made this technique inaccurate if only one TLD reading had been made. However, the ratio of the TLD doses from the cylindrical targets to those from the solid anodes was calculated to be independent of voltage, to within the accuracy of the Monte Carlo calculation, over the range 0.5–1.6 MeV. This ratio was used to determine the transport efficiency for the 3.2- and 1.9-cm diameter cylinders.

The presence of electron drift losses near the ends of the cylinders would have complicated interpretation of TLD data for the smaller diameter cylinders. However, the X-ray pinhole camera was able to resolve this loss spatially and thus was used to infer the transport efficiencies to smaller cylinders. A typical X-ray photograph is shown in Fig. 2. Previous work¹⁶ has shown that the image on the developed film is proportional to the energy absorbed in the film emulsion in this regime, and absolute calibrations were not required. The spatial distributions in the X-ray images were in good agreement with the theoretical predictions. The observed X-ray intensities were compared with the predicted intensities calculated using the techniques described above. By normalizing the experimental photographic intensities to the theory for the 3.2-cm diameter

cylinder, the absolute transport efficiencies could be inferred for smaller diameter cylinders.

Figure 3 compares the transport efficiencies predicted using the numerical model and the experimental data. Surprisingly, the agreement between theory and experiment is much better than possible inaccuracies predict. The only significant disagreement occurs for the smallest cylinder for which the measurement is $\sim 5\%$ below the calculated transport efficiency of 19%. This result is strong support for the models used to calculate plasma channel formation, channel overlap geometry, electron and photon transport, and for the experimental procedures used. Figure 3 also shows a calculated prediction of transport efficiency for protons injected into the same channel system used here for the electron beams. The ions would reach a smaller radius with higher efficiency than do the electrons, and thus would irradiate a target with higher power. This result, along with their more favourable energy deposition properties³ and the recent rapid advances in ion diode research^{1,4}, makes light ions more attractive than electrons for ICF use.

We thank P. Davis, D. Noack, R. Noack, W. Simpson and D. Zagar for technical assistance, and the data acquisition and Proto II operations crews for their perseverance and help. This work was supported by the US Department of Energy under Contract DE-AC04-76-DP000789.

Received 24 March; accepted 9 May 1980.

1. Miller, P. A., Johnson, D. J., Wright, T. P. & Kuswa, G. W. *Comments Plasma Phys. and Controlled Thermonucl. Fus.* **5**, 95–103 (1979).
2. Widner, M. M. *et al. Phys. Rev. Lett.* **43**, 357–360 (1979).
3. Clauser, M. J. *Phys. Rev. Lett.* **35**, 848–851 (1976).
4. Johnson, D. J. *et al. Phys. Rev. Lett.* **42**, 610–613 (1979).
5. Wright, T. P. *J. appl. Phys.* **49**, 3842–3850 (1978).
6. Wright, T. P. & Halbleib, J. A. Sr, *Sandia Rep. SAND79-2064J*; *Phys. Fluids* (in the press).
7. Halbleib, J. A. Sr & Wright, T. P. *Sandia Rep. SAND79-2063J*; *Phys. Fluids* (in the press).
8. Miller, P. A. *et al. Phys. Rev. Lett.* **39**, 92–95 (1977).
9. Benford, J. *J. appl. Phys.* **48**, 2320–2323 (1977).
10. Ottinger, P. F., Mosher, D. & Goldstein, S. A. *Phys. Fluids* **22**, 332–337 (1979).
11. Mosher, D., Colombant, D. G., Goldstein, S. A. & Lee, R. *IEEE Conf. Record-Abstr. 1978 IEEE Int. Conf. Plasma Sci.* 113 (IEEE, New York, 1978).
12. Freeman, J. R. *et al. IEEE Conf. Record-Abstr. 1978 IEEE Int. Conf. Plasma Sci.* 177 (IEEE, New York, 1978).
13. Martin, T. H., Johnson, D. L. & McDaniel, D. H. *Proc. 2nd Int. Topical Conf. High Power Electron and Ion Beam Research and Technology Vol. 2*, 807–818 (Cornell University, New York, 1977).
14. Halbleib, J. A. Sr & VanDevender, W. H. *J. appl. Phys.* **48**, 2312–2319 (1977).
15. Wright, T. P. & Halbleib, J. A. Sr *Bull. Am. phys. Soc.* **24**, 949 (1979).
16. Dozier, C. M., Brown, D. B., Birks, L. S., Lyons, P. B. & Benjamin, R. F. *J. appl. Phys.* **47**, 3732–3739 (1976).

Opto-acoustic study of weak optical absorption of liquid methane

C. K. N. Patel, E. T. Nelson & R. J. Kerl

Bell Laboratories, Murray Hill, New Jersey 07974

The first opto-acoustic absorption spectra of liquid methane at 94 K in the region of 5,500–7,500 Å is reported here. The observed absorption bands are very weak with peak absorption coefficients of 16×10^{-3} to $1 \times 10^{-4} \text{ cm}^{-1}$ and are of importance in the study of spectra of outer planets (such as Jupiter and Neptune) for detecting the existence of liquid methane on these planets.

Methane is a minor constituent of the atmospheres of the outer major planets¹, Uranus, Neptune, Jupiter and Saturn, as well as of Titan, the largest satellite of Saturn. This is deduced from the absorption bands of CH_4 which dominate the visible and near IR spectra of these planets². The long effective optical path lengths in the atmosphere of these planetary bodies enables the absorption bands to be readily observed using ground based telescopes. Further, the temperatures on some of these planetary bodies are such that CH_4 could exist as liquid droplets³ or even as an ocean⁴ on Titan. To provide correlations with laboratory data, and to deduce the CH_4 concentration of the atmospheres of the above planets from the optical astronomy data, an accurate determination of the positions of these weak

absorption bands, their band strengths, and temperature dependences is needed. Because CH_4 at pressures of tens of atmospheres or even at liquid densities (~ 590 amagats) is very transparent in the visible region the laboratory data are difficult to obtain. Conventional optical absorption spectroscopy in the $\sim 4,000$ – $11,000$ Å region has recently been carried out using a multipass cell to obtain absorption pathlengths of ~ 1 km for room temperature CH_4 at pressures up to 20 atm (ref. 5) resulting in density-pathlength products of up to 20 km amagat. Longer path lengths of up to 3.86 km at lower pressures were used in earlier studies^{16,17}. These data show several absorption bands of generally decreasing strength towards shorter wavelengths. Conventional optical spectroscopy of liquid CH_4 has been recently reported⁶ in the region of 6,190–19,000 Å using a cell length of ~ 3 cm. The sensitivity of these studies, however, is such that the shortest wavelength absorption feature reported is at $\lambda \sim 6,190$ Å with a peak absorption coefficient of $\sim 4 \times 10^{-3} \text{ cm}^{-1}$. The signal-to-noise ratio reported for this peak is only about 5:1, that is, the 6,190 Å peak height is ~ 5 times the system noise. If we extrapolate from Giver's high temperature gas data⁵, liquid CH_4 should exhibit weaker absorption bands all throughout the visible region (4,000–7,000 Å). Measurements of these absorption bands are clearly beyond the capabilities of the conventional spectroscopy studies of Ramaprasad *et al.*⁶ who used a density path length product ~ 19 m amagat although some improvement is possible by using longer (but perhaps experimentally difficult) path lengths. Note that the multipass long cell technique used by Giver⁵ is not applicable to liquid CH_4 studies because of obvious practical problems.

We report here optical absorption spectrum of liquid CH_4 at ~ 94 K in the 5,500–7,500 Å region. Several absorption bands are observed, the strongest one being at $13,700 \text{ cm}^{-1}$ with an absorption coefficient of $(1.6 \pm 0.2) \times 10^{-2} \text{ cm}^{-1}$ and the weakest one at $15,515 \text{ cm}^{-1}$ with an absorption coefficient of $\sim (1 \pm 0.3) \times 10^{-4} \text{ cm}^{-1}$. The signal-to-noise ratio for the weakest peak reported is such that we should be able to observe absorption features as small as $\sim 10^{-6} \text{ cm}^{-1}$. Our measurements were carried out using an opto-acoustic (OA) detection technique⁷ which has been shown to be ideally suited for measuring absorption coefficients of room temperature liquids at levels as low as $\sim 10^{-7} \text{ cm}^{-1}$ and which we have adapted for the study of cryogenic liquids. The OA cell used in the present liquid CH_4 measurements had an effective density path length product of ~ 6.33 m amagat. In addition to the position and the strengths of the absorption bands, we report their bandwidths, asymmetries and identifications. In general, the absorption peaks arise from combinations of harmonics of the various vibrational fundamentals of CH_4 . Further, the absorption bands for liquid CH_4 show asymmetries and frequency shifts (from the gas phase data) which are reminiscent of the liquid benzene results⁸.

The measurements were carried out using a pulsed laser immersed transducer OA gated detection technique⁷, modified for low temperature studies. The all stainless steel OA cell⁹ was attached to a liquid N_2 cryostat such that any desired temperature between 80 and 200 K can be obtained. Methane (Matheson, Ultra-High Purity grade) was directly condensed into the low temperature cell and its pressure was monitored using a capacitance manometer. Tunable pulsed laser radiation was obtained from either a flashlamp pumped dye laser (pulse energy ~ 1 – 3 mJ, pulse duration $\tau_p \sim 1$ μs , and repetition rate ~ 10 pulses per s (p.p.s.)), or a doubled Nd:YAG pumped dye laser (pulse energy ~ 1 – 5 mJ, $\tau_p \sim 10$ ns, and pulse repetition rate ~ 20 p.p.s.). The laser pulse energy was measured using a Laser Energetics pulsed joule-meter. The OA signal as well as the reference laser level signal sampled by box-car integrators were averaged and ratioed by a NOVA 800 minicomputer. We thus obtained spectra of normalized OA signal $S_r = V/E_{\text{inc}}$, where V is the OA signal from the PZT transducer and E_{inc} is the incident laser pulse energy, as a function of the dye laser frequency.

The transient acoustical output signal from the transducer (in form of ringing due to the 200 kHz resonant frequency of the

Table 1 Observed data on optical absorption spectrum of liquid methane

Feature‡	Bandcentre§ (cm ⁻¹)	Liquid*		Asymmetry¶	Identification	Gas† Bandcentre (cm ⁻¹)
		Peak absorption (cm ⁻¹)	Width (cm ⁻¹)			
1	13,700	1.6 × 10 ⁻²	180	1.69	4ν ₁ + ν ₃	13,755
2	14,170	1 × 10 ⁻³	—	—	3ν ₁ + 2ν ₃	14,225
3	14,585	3 × 10 ⁻⁴	—	—	2ν ₁ + 3ν ₃	14,652
4	14,935	7 × 10 ⁻⁴	250	2.4	4ν ₁ + ν ₃ + (ν ₂ , ν ₄)	14,992
5	15,170	—	—	—	—	—
6	15,300	—	—	—	—	—
7	15,410	—	—	—	3ν ₁ + 2ν ₃ + (ν ₂ , ν ₄)	—
8	15,515	1.2 × 10 ⁻⁴	—	—	—	15,576
9	16,090	3 × 10 ⁻³	224	1.67	5ν ₁ + ν ₃	16,155
10	16,714	2.3 × 10 ⁻⁴	240	—	4ν ₁ + 2ν ₃	16,750
11	17,320	2.5 × 10 ⁻⁴	310	1.38	5ν ₁ + ν ₃ + (ν ₂ , ν ₄)	17,361

* T ≈ 94 K, density length product 6.3 M amagat.

† T ≈ 300 K, data from ref. 5.

‡ See Fig. 1.

§ Bandcentres generally accurate to ±5 cm⁻¹ for strong bands.

|| Accuracy ±30%.

¶ See text for definition.

PZT transducer, and the reflections of acoustic waves generated from the condensed liquid) is delayed from the incident laser pulse by a delay, $\tau_D = d/v_a$, where d is the distance between the laser illuminated volume and the OA transducer, and v_a is the acoustic velocity in the condensed liquid. From a measurement of τ_D versus d we have determined the sound velocity in liquid CH₄ ($T = 94$ K) to be $v_a = (1,455 \pm 18) \text{ m s}^{-1}$. This value is consistent with the currently accepted value¹⁰.

Figure 1 shows the measured spectrum of normalized OA signal from liquid CH₄ at 94 K from 13,300 to 17,800 cm⁻¹. Several clearly recognizable absorption peaks are indicated. Spectral resolution which is typically ~1.5–2 cm⁻¹ is much smaller than the observed widths of the absorption features. Note that the only previous measurements of liquid CH₄ in the visible reported⁶ the strong peak at ~16,090 cm⁻¹ with a signal-to-noise ratio of ≈ 5. We see many more absorption features with peak absorptions which are ~30 smaller than that for the 16,090 cm⁻¹ band. And even for the weak features the signal-to-noise ratio exceeds 100 for an integration time of ~1.5 s per resolution element of 1.5–2 cm⁻¹. Note also that there are no regions of spectrum where liquid CH₄ absorption drops below ~10⁻⁴ cm⁻¹. All the observed absorption features show a pronounced asymmetry, the absorption tails being longer on the high frequency side of the peaks.

As reported earlier^{7,9}, the normalized OA signal, S_v , can be related to the intrinsic absorption coefficient of the condensed material through the relationship

$$S_v = \frac{V}{E_{inc}} = K\beta \frac{v_a}{C_p} \alpha_v \quad (1)$$

where β is the thermal volumetric expansion coefficient, C_p the specific heat at constant pressure, α_v the absorption coefficient ($\alpha_v l \ll 1$ where l is the sample path length), and K is a constant that describes the OA conversion efficiency and includes such parameters as the geometry of the OA cell, the acoustic impedance mismatch between the liquid and the transducer, and the sensitivity of the PZT transducer. Thus a knowledge of all the relevant parameters in equation (1) is sufficient to extract the absorption coefficient from the observed normalized OA signal. Procedures such as this have provided an excellent estimate of the absorption coefficient of media at room temperature where a sensitivity $\Sigma = S_v/\alpha_v$, calibrated for one liquid can be approximated for another of known β , C_p , and ρ . An accurate determination of the sensitivity Σ for a given liquid requires that a suitable dye be incorporated into the liquid medium such that a very small concentration of the dye (small enough so that none of the parameters of the right-hand side of equation (1) other than α_v are perturbed) produces an absorption which can be measured accurately by conventional optical techniques (and by

the OA technique). Unfortunately no data are available for the low temperature condensed gases. For the present study, we have extrapolated our sensitivity measurements from benzene studies^{7,8} to obtain absorption coefficient from the measured S_v for liquid CH₄. For the strong absorption peak at 16,090 cm⁻¹, we thus obtain $\alpha(16,090) \approx (3 \pm 1) \times 10^{-3} \text{ cm}^{-1}$. This value is in a reasonable agreement with the one which can be deduced from ref. 6.

Table 1 shows the data on observed absorption bands including: (1) the band peak frequencies; (2) bandwidths of peaks which are sufficiently separated to make such a measurement meaningful; (3) band asymmetry parameter defined as the ratio of the half-width at half-height on the high frequency side of peak to that on the low frequency side of the peak; (4) the peak absorption coefficient (subtracting out the broad background absorption in case of weak absorption features); and (5) identifications of the peaks, using Giver's⁵ scheme for gaseous methane. We also show in Table 1 relevant data from the gas phase studies of Giver⁵.

(1) There is a systematic shift of ~35–50 cm⁻¹ of the liquid CH₄ absorption peak positions towards low frequency side when compared with gaseous CH₄ data. This shift is thought to arise from the effect of local environment in liquid CH₄ and is qualitatively in agreement with the data on liquid⁸ and gaseous benzene (ref. 11 and R. G. Bray, personal communication). Note that our liquid CH₄ data show well-defined subsidiary peaks/shoulders on the absorption features which are not visible in the gas data of Giver⁵.

(2) The observed liquid CH₄ bandwidths vary from ~250 to 350 cm⁻¹ and are much wider than those for gaseous CH₄, again in qualitative agreement with liquid/gaseous benzene studies. The higher harmonics in a given series have larger bandwidths.

(3) All the observed liquid CH₄ absorption features in the 5,500–7,500 Å region show pronounced asymmetries with the asymmetry parameter varying from ~1.4 to 2.4. The gaseous CH₄ (ref. 5) data on the other hand show no such asymmetry

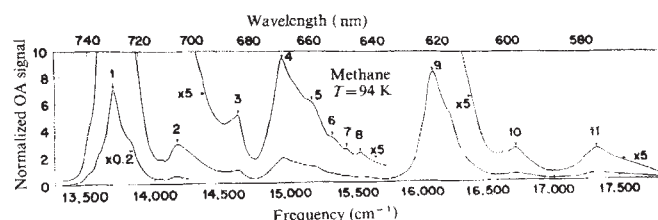


Fig. 1 Normalized OA signal as a function of dye laser frequency. Data including the absolute absorption coefficients for the indicated peaks are summarized in Table 1.

which is again thought to arise from local environment interactions in qualitative agreement with liquid/gaseous benzene studies (refs 8, 11 and R. G. Bray, personal communication) of its harmonic absorption in the visible region. We observe that the asymmetry of the absorption is a general characteristic feature of the absorption profiles of liquids.

(4) From the measured bandshapes one can derive integrated band strength (not done here). The absorption coefficients of the measured peaks vary from $\sim 1 \times 10^{-4}$ to $\sim 1.6 \times 10^{-2} \text{ cm}^{-1}$. From the signal-to-noise ratio for the smallest peaks (obtained using an integration time of 1.5 s per resolution element of 1.5 cm^{-1}), we conclude that we have the ability to measure absorption features as small as $\sim 10^{-6} \text{ cm}^{-1}$ for liquids at low temperatures. The trend of the measured absorption coefficient for liquid CH_4 is similar to that for gaseous CH_4 , that is the higher the harmonics the smaller the absorption coefficients.

(5) While major peaks have been reasonably well identified as arising from the combinations of harmonics of the four fundamental vibrational frequencies of CH_4 , the minor peaks are not yet assigned.

The detailed knowledge about liquid CH_4 absorption spectrum provided here should now allow us to look carefully at the spectra of the outer planets to search for existence of liquid CH_4 . (As pointed out by Ramaprasad *et al.*⁶, detection of liquid CH_4 , in form of an ocean on the planetary surface, is likely to be very difficult because only surface reflection is involved. However, for liquid CH_4 droplets in the planetary atmosphere, the sunlight penetrates and is reflected back from a surface or a cloud layer. Thus detection or setting an upper limit of the droplet fraction of CH_4 is feasible if high resolution high quality astronomical spectroscopic data are available.) Important information is to be gained from the position of the absorption bands as well as from their widths and shapes. As is seen above, these three features are characteristically different for liquid CH_4 from those for gaseous CH_4 . For example, the existence of liquid CH_4 (in the form of droplets in the planetary atmosphere) can be deduced

from the planetary astronomical absorption data if measurable differences are found in the peaks (positions, shapes, and widths) of the absorption bands (compared to the absorption peaks of gaseous CH_4).

Finally, our demonstration of the usefulness of liquid OA absorption measurement technique to low temperature liquids opens up an entirely new field of study. The ability to measure absorption coefficients as small as 10^{-6} cm^{-1} (an ability which we feel can be improved considerably) at cryogenic temperatures will undoubtedly provide exciting new data on liquids and solids. We have already begun to look at other liquified gases such as ethylene, and ammonia which are also important¹²⁻¹⁴ in the atmospheres of the outer planets. We are modifying our OA measurement techniques to go to yet lower temperatures $\sim 4 \text{ K}$ to investigate H_2 and HD which are the major constituents¹⁵ of these planets.

We thank Drs L. J. Lanzerotti and J. A. Tyson for helpful comments.

Received 21 March; accepted 20 May 1980.

1. Kuiper, G. P. (ed.), *The Atmospheres of Earth and Planets* (University of Chicago Press, 1952).
2. Lutz, B. L., Owen, T. & Cess, R. D. *Astrophys. J.* **203**, 541 (1976).
3. Danielson, R. E. *Icarus* **30**, 462 (1977).
4. Caldwell, J. in *Planetary Satellites* (ed. Burns, J. A.) 438-450 (University of Arizona Press, Tucson, 1977).
5. Giver, L. P. *J. quant. Spectrosc. Radiat. Transfer* **19**, 311 (1978).
6. Ramaprasad, K. R., Caldwell, J. & McClure, D. S. *Icarus* **35**, 400 (1978).
7. Patel, C. K. N. & Tam, A. C. *Appl. Phys. Lett.* **34**, 467 (1979).
8. Patel, C. K. N., Tam, A. C. & Kerl, R. J. *J. chem. Phys.* **71**, 1470 (1979).
9. Patel, C. K. N. & Tam, A. C. *Nature* **280**, 302 (1979); *Appl. Opt.* **18**, 3348 (1979).
10. *International Thermodynamic Tables of Fluid State* Vol. 5 (Pergamon, Oxford, 1976).
11. Reddy, K. V., Bray, R. G. & Berry, M. J. in *Advances in Laser Chemistry* (ed. Zewail, A.) 48 (Springer, Berlin, 1978).
12. Ridgeway, S. T. *Astrophys. J. Lett.* **187**, L41 (1974).
13. Combes, M., Encrenaz, T., Vapillon, L., Zeau, Y. & Lesqueren, C. *Astr. Astrophys.* **34**, 33 (1974).
14. Encrenaz, T., Owen, T. & Woodman, J. H. *Astr. Astrophys.* **37**, 49 (1974).
15. McKellar, A. R. W. & Welsh, H. L. *Proc. R. Soc. A* **322**, 421 (1971).
16. Dick, K. A. & Fink, U. *J. quant. Spectrosc. Radiat. Transfer* **18**, 433 (1977).
17. Fink, U., Benner, D. C. & Dick, K. A. *J. quant. Spectrosc. Radiat. Transfer* **18**, 447 (1977).

Contact angles associated with thin liquid films in emulsions

M. P. Aronson & H. M. Princen

Lever Brothers Company, Research Division, 45 River Road, Edgewater, New Jersey 07020

As a result of interaction forces, the tension of a thin liquid film can be lower than the sum of the interfacial tensions of the bulk surfaces bounding the film¹. To satisfy conditions of mechanical equilibrium, this reduced film tension leads to a macroscopic contact angle between the film and its associated Plateau border². We have observed and report here very high contact angles, in some cases approaching 90° , for aqueous films between oil droplets, stabilized by anionic surfactants in the presence of various electrolytes. The contact angle is found to depend on the nature of the surfactant head group, type and concentration of counter-ion, and temperature. These contact angles have a pronounced effect on the stability, structure and rheology of emulsions.

Consider the thin metastable liquid film and its Plateau border that form when a vertical rectangular glass frame is withdrawn from a surfactant solution (Fig. 1). The tension of the film, γ_t , is related to the normal bulk interfacial tension, γ by³

$$\gamma_t = 2\gamma + \Delta F = 2\gamma + \int_0^{h_e} \pi(h) dh \quad (1)$$

h is the film thickness, h_e is the equilibrium thickness, π is the 'disjoining pressure', and ΔF is the free energy of interaction per unit area of the film, that is, the reversible work of thinning the

film from infinite thickness, where the surfaces do not interact, down to the equilibrium thickness, where attractive and repulsive interaction forces are in balance. [Strictly speaking³, equation (1) should include a term πh_e . For the extremely thin films (so-called Newton black films) to be discussed in this study, this term is expected to be negligible.]

A mechanical analysis⁴ of the region where the film and the Plateau border join indicates that, whenever $\Delta F < 0$, the two bulk surfaces will, on a macroscopic scale, intersect at a sharp angle, 2θ , such that

$$\gamma_t = 2\gamma \cos \theta \quad (2)$$

It follows from equations (1) and (2) that

$$-\Delta F = 2\gamma(1 - \cos \theta) \quad (3)$$

One of the many manifestations of a finite contact angle is the spontaneous clumping that occurs when two or more oil drops approach each other in an aqueous surfactant solution (Fig. 2).

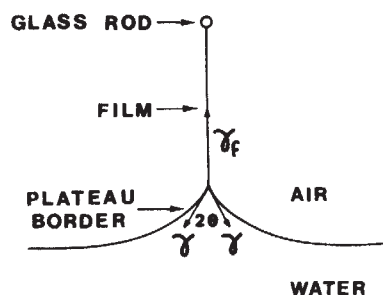


Fig. 1 Film and associated plateau border that are formed when a vertical glass frame is pulled out of a surfactant solution.

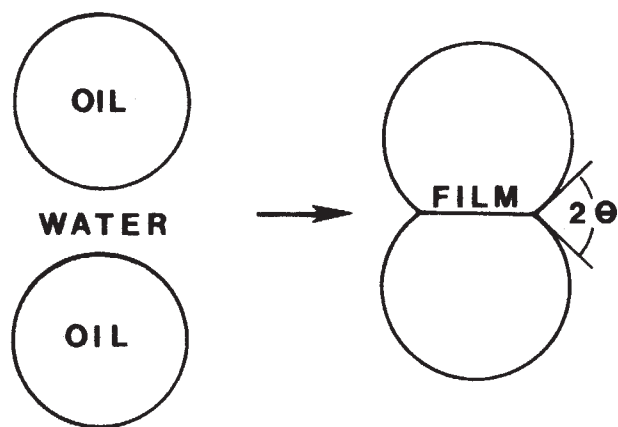


Fig. 2 The spontaneous flocculation of oil droplets when the contact angle, θ , is greater than zero.

When the contact angle is large, the droplets comprising the cluster are highly deformed, while remaining separated by an extremely thin film of intervening liquid of reduced tension.

Appreciable contact angles, as high as 15° , have been observed in aqueous films in air⁵⁻⁸. For lipid films in water, contact angles up to 8° have been reported^{9,10}. If one assumes that interaction free energies for aqueous films in oil are of about the same magnitude as for aqueous films in air, one would, on the basis of equation (3), expect much larger angles for the former on account of their generally much lower interfacial tension.

Indeed, while studying¹¹ concentrated oil-in-water emulsions, we have observed very large contact angles, in some cases approaching 90° which would seem to be the theoretical limit. The emulsions were stabilized with anionic surfactants and contained added electrolyte. The contact angle was found to depend strongly on the composition of the aqueous phase (type of surfactant, and concentration and kind of counter-ion), while the nature of the oil phase had a relatively minor influence. We have used two classes of anionic surfactant: sodium alkyl sulphates and sodium alkyl carboxylates (soaps) in the presence of added LiCl, NaCl or KCl. With alkyl sulphates, the contact angle reached at a given electrolyte concentration increases in the order $\text{Li}^+ \ll \text{Na}^+ < \text{K}^+$. With alkyl carboxylates, the order seems to be reversed, although there is some doubt about the position of Li^+ because of the very low solubility of lithium soaps. Similar reversals in the relative effect of the alkali metal ions on properties of alkyl carboxylates and alkyl sulphates have been reported in other areas of colloid and surface chemistry¹². With Na^+ as the counter-ion, the alkyl carboxylates give rise to much higher contact angles than do the alkyl sulphates. For example, with 0.3 M NaCl and 0.2% surfactant at 25°C , the contact angles with dodecane are $\approx 0^\circ$ for SDS and $\approx 57^\circ$ for sodium oleate. Contact angles were also observed with other *n*-alkanes, as well as with paraffin oil and carbon tetrachloride. The results of this study, still in progress, will be presented elsewhere.

More recently, we have observed that the contact angle varies strongly with temperature. This is not unexpected, as this effect has been reported also for aqueous films in air⁸. Because of the much greater range of contact angles that can be obtained with aqueous films between oil, however, the phenomenon is much more dramatic. An example is given in Fig. 3 that shows dodecane droplets in an aqueous phase containing 0.25% sodium oleate, 0.25% sodium laurate and 0.6 M NaCl. At 25°C , this system has a contact angle of about 70° which leads to the formation of flocs composed of highly deformed oil droplets. The floc is quite rigid and the individual droplets do not easily roll past each other. Very little water is seen to remain in the Plateau borders in the interior of each floc. Just as in polyhedral foams, these borders are of two types: 'linear' channels where three films meet at angles of 120° , and 'tetrahedral' corner pockets where four linear borders meet at angles of 109.47° . It

can be shown¹¹ that the linear Plateau borders empty out when $\theta \geq 30^\circ$ while the tetrahedral pockets do so when $\theta \geq 35.26^\circ$.

When the dispersion is heated, the contact angle decreases and the cluster gradually decomposes into increasingly spherical droplets that begin to shift and rearrange (45°C). Finally, at 60°C , the contact angle is near 0° , and the floc now appears as a collection of discrete spherical droplets that are mobile. If, at this stage, the dispersion is cooled back to 25°C , rigid flocs of deformed droplets again appear. They do not, of course, revert to the original configuration because of shifts that have taken place in the first cycle.

The presence of large contact angles can have a pronounced effect on several properties of an emulsion. For example, there is a dramatic increase in the rate and extent of sedimentation or creaming of an emulsion when $\theta > 10^\circ$.

The structure of a concentrated emulsion, such as the cream layer formed on top of an initially dilute emulsion, is also very sensitive to the contact angle^{11,13}. When the cream is formed by

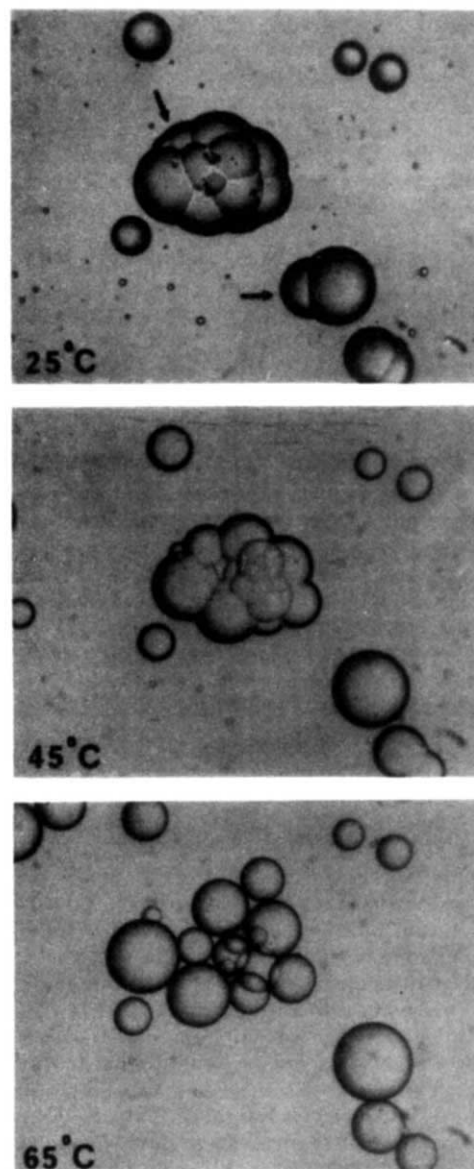


Fig. 3 The effect of temperature on the contact angle in flocs formed when dodecane droplets are dispersed in an aqueous solution of 0.25% sodium oleate, 0.25% sodium laurate and 0.6 M NaCl. The arrows indicate droplets that disappear during the experiment because of coalescence. Magnification, $\sim 90\times$. An Olympus MG inverted microscope was used and heating was accomplished with a Mettler FP2 programmable hot stage. Contact angle equilibrium was generally achieved in less than 10 min.

sedimentation under gravity, an open network of flocs can form when $\theta \gg 0$. Appreciable amounts of water can be trapped in this network, leading to an expanded cream whose volume fraction of oil is < 0.74 (the volume fraction corresponding to closest packing of monodisperse spheres). In this regard, this type of emulsion cream is analogous to the voluminous sediment encountered with a highly unstable colloidal dispersion of solid particles¹⁴. When such an expanded emulsion cream is compacted by mild centrifugation and then allowed to relax and equilibrate under normal gravity, much of the occluded water can be squeezed out, so that the resulting cream tends to have a final volume fraction of oil significantly greater than 0.74, as predicted by theory for the state of minimum free energy^{11,13}. Indeed, using this method, we have prepared stable creams whose volume fraction is as high as 0.94 (ref. 11).

Note that some aspects of this work may have their parallel in biological systems. The cluster shown in Fig. 3 has a close morphological resemblance to a blastula, that is, the aggregate of living cells at the early stage of embryonic development. Of course, in biological systems the liquid phases would be reversed, that is, aqueous drops (cells) would be separated by lipid films (membranes).

We thank Lever Brothers Company for permission to publish this work.

Received 8 April; accepted 23 May 1980.

- Derjaguin, B. V. *Acta physicochim. URSS* **12**, 181–200 (1940).
- Princen, H. M. & Mason, S. G. *J. Colloid Interface Sci.* **29**, 156–172 (1965).
- De Feijter, J. A., Rijnbout, J. B. & Vrij, A. *J. Colloid Interface Sci.* **64**, 258–268 (1978).
- De Feijter, J. A. & Vrij, A. *J. electroanal. Chem.* **37**, 9–22 (1972).
- Princen, H. M. *J. phys. Chem.* **72**, 3342–3345 (1968).
- Huisman, F. & Mysels, K. J. *J. phys. Chem.* **73**, 489–497 (1969).
- Kolarov, T., Scheludko, A. & Exerowa, D. *Trans. Faraday Soc.* **64**, 2864–2873 (1968).
- De Feijter, J. A. & Vrij, A. *J. Colloid Interface Sci.* **64**, 269–277 (1978).
- Haydon, D. A. & Taylor, J. L. *Nature* **217**, 739–740 (1968).
- Requena, J., Billett, D. F. & Haydon, D. A. *Proc. R. Soc. A* **347**, 141–159 (1975).
- Princen, H. M., Aronson, M. P. & Moser, J. C. *J. Colloid Interface Sci.* **75**, 246–270 (1980).
- Goddard, E. D. *Croatia chem. Acta* **42**, 143–150 (1970).
- Princen, H. M. *J. Colloid Interface Sci.* **71**, 55–66 (1979).
- Vold, M. J. *J. Colloid Interface Sci.* **18**, 684–695 (1963).

Hydrostatic pressure and long-chain organic molecule diffusion through a polymer matrix

A. R. Rennie & D. Tabor

Department of Physics, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

Earlier papers from our laboratory^{1–3} have described a sensitive but relatively simple technique using IR spectroscopy to study the diffusion of labelled organic molecules through a polymer matrix. The method showed that the behaviour agreed well with de Gennes⁴ model of reptation which assumes that long organic molecules diffuse by the motion of segments consisting of 20 to 30 units in the backbone. We describe here an extension of this work and deal with the effect of hydrostatic pressure on the diffusion of two hydrocarbons (suitably labelled) with backbone units $N = 18$ and 45 through molten low density branched polyethylene at pressures ranging from atmospheric to 1,300 atmospheres. The results show that the diffusion coefficient D decreases as the pressure P is increased. To a close approximation $\log D$ decreases linearly with P implying an apparent activation volume, V . Typical values for $N = 18$, are $V = 55 \text{ \AA}^3$ at 150°C and $70 \text{ \AA}^3$ at 120°C implying a thermal expansion of the free volume of order $0.5 \text{ \AA}^3 \text{ K}^{-1}$. For $N = 45$, $V = 80 \text{ \AA}^3$ at 120°C .

The experiments were carried out using ICI Alkathene 11/01 branched polyethylene as the matrix. The diffusants were stearamide ($\text{CH}_3(\text{CH}_2)_{16}\text{CONH}_2$) with 18 backbone atoms and behenyl behenate ($\text{CH}_3(\text{CH}_2)_{20}\text{COOCH}(\text{CH}_2)_{20}\text{CH}_3$) with 45 units counting the oxygen in the ester linkage. Cylindrical

specimens were prepared in the manner described earlier¹ one half containing about 1% diffusant and the other consisting of matrix polymer only. A thin slice cut from the cylinder perpendicular to the join is scanned in an IR microdensitometer tuned to the adsorption frequency of the carbonyl group in the diffusant. Initially the concentration is a step function. As a result of diffusion the profile changes and from the new shape the diffusion coefficient can be calculated.

The diffusion experiments are all carried out above the melting point of the polymer. First, this avoids problems involving different degrees of crystallinity as the molten material at any specified temperature has a reproducible 'morphology'. Second, the diffusion rates are far more rapid than with the solid polymer, an important practical consideration in view of the decrease in D as the pressure is increased. Third, the specimen may be contained in a metal cylinder and compressed by a piston—because it is liquid the pressure is entirely hydrostatic. The force on the piston is applied with dead weights using a converted lever-arm tensile-testing machine and in this way the pressure is maintained constant without the need for elaborate servo-mechanisms. The seals, which slide in the cylinder, are of the Bridgman unsupported area type⁵. The metal cylinder is heated by electrical resistances and thermocouples are inserted at various points near the specimen. Temperature control is reproducible to $\pm 1^\circ\text{C}$. After the specimen has been subjected to the required pressure and temperature for a specified time it is solidified rapidly, removed from the containing cylinder and then sliced for examination in the IR microdensitometer.

Figure 1 shows that $\log D$ decreases linearly with P . Assuming a simple Arrhenius relation of the form $D = D_0 \exp(-PV/kT)$ where V is the activation volume for diffusion, k the Boltzmann constant and T the absolute temperature, we obtained the values quoted in Table 1.

From Table 1 we first note that the activation volumes are very much smaller than the molecular volumes of the diffusants themselves ($400 \text{ \AA}^3$ for stearamide, $1,000 \text{ \AA}^3$ for behenyl behenate). It is not possible to provide a specific model to explain this. It could be associated with a crankshaft motion involving some 3–4 CH_2 units; or it could relate to a small overall dilation of the matrix to allow passage of the diffusant. However, we also note that the activation volume for the behenyl behenate is only a little larger than for the stearamide although the molecular volume is more than twice as large. This is consistent with the earlier work of Klein and Briscoe^{2,3} which

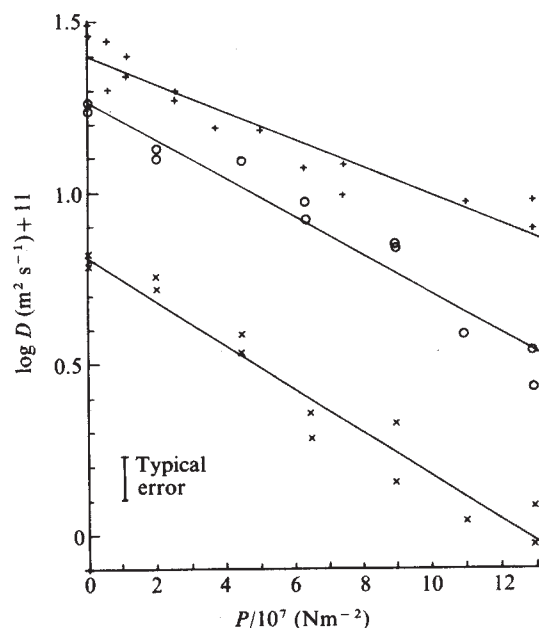


Fig. 1 Plot of $\log D$ against applied hydrostatic pressure, P , for stearamide at 150°C (+) and 120°C (○) and behenyl behenate at 120°C (×).

Table 1 Activation volumes, V , for diffusants in molten branched polyethylene

Diffusant	N	Temperature (°C)	V (Å ³)
Stearamide	18	150	55 ± 7
Stearamide	18	120	70 ± 8
Behenyl behenate	45	120	80 ± 8

indicated that however long the diffusant, motion occurs by the movement within individual segments each of which is about 20 carbons in length. Whatever the detailed mechanism it is clear that the diffusant is constrained by the surrounding matrix so that the activation process is easier the larger the available free volume. This accounts for the observation that the activation volume for stearamide decreases from 70 Å³ at 120 °C to 55 Å³ at 150 °C. Indeed, we may attribute this decrease to a thermal expansion of the free volume of the order 0.5 Å³ K⁻¹.

It is interesting to compare these values for V with those deduced from the shear properties of polyethylene in the solid state. Ward⁶ and Wilding and Ward⁷ quote a value of ~50 Å³ for bulk shear of linear polyethylene. Although the polymer was largely crystalline and highly oriented the deformation process

involves radical disorganization of both the crystalline and the amorphous material in the specimen. Consequently, the mechanisms of motion may not be greatly different from those occurring in the melt. Similarly B. J. Briscoe (personal communication) quotes a value between 100 and 200 Å³ for the shear of thin films of branched polyethylene. In these experiments the molecular length (and hence N) was many hundreds of times greater than the molecular length of the diffusants used in the present work. This again suggests that the unit involved in shear is a short segment, independent of the overall length of the polymer molecule. We hope to explore this aspect further by carrying out diffusion experiments with diffusants of increasing molecular length. If these data are to be used in the construction of molecular models of diffusion it should be noted once again that the activation volumes are very much smaller than the molecular volume of the segment itself.

A.R.R. thanks the SRC and ICI for a CASE studentship.

Received 31 March; accepted 30 May 1980.

1. Klein, J. & Briscoe, B. J. *Nature* **257**, 386 (1975); *Polymer* **17**, 481 (1976).
2. Klein, J. *Nature* **271**, 143 (1978).
3. Klein, J. & Briscoe, B. J. *Proc. R. Soc. A* **365** (1979).
4. de Gennes, P. G. *J. Chem. Phys.* **55**, 572 (1971).
5. Bridgman, P. W. *The Physics of High Pressure* 39 (Bell, London, 1949).
6. Ward, I. M. *Phil. Trans. R. Soc. A* **294**, 473 (1980).
7. Wilding, M. A. & Ward, I. M. *Polymer* **19**, 969 (1978).

Possible ozone reductions and UV changes at the Earth's surface

J. A. Pyle

Department of Atmospheric Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3BD, UK

R. G. Derwent

Environmental and Medical Sciences Division, AERE Harwell, Oxfordshire OX11 0RA, UK

During recent years there have been suggestions¹⁻³ that man's activities might reduce the thickness of the atmospheric ozone layer. Reductions in ozone could then give rise to an increase in potentially harmful solar UV radiation at the Earth's surface. Concern has been expressed particularly about the possibility of subsequent increases in skin cancer. Most calculations of ozone reduction have employed one-dimensional (height-only) models and few have calculated the solar radiation reaching the ground. Rather, it has been assumed that, for example, a 1% depletion in ozone is equivalent to a 2% increase in UV-B (the biologically important radiation in the wavelength range 280–320 nm) (refs 4, 5). However, studies using two-dimensional models^{6,7} indicate that any change in ozone will not be uniform, but will show significant variations with latitude and season. We show here first, that the commonly accepted idea that a depletion of ozone of a particular figure is equivalent to an UV increase of twice that figure does not hold generally; second, we point out the importance of seasonal and latitudinal variations in ozone changes on the UV radiation reaching the Earth's surface.

Figure 1 shows the latitude-time cross-section of the depletion in total ozone calculated by one of us (J.A.P.) using a two-dimensional model of the atmosphere including detailed photochemistry and simulating a source of chlorofluorocarbons (CFCs). The features of the depletion are similar to a calculation in the model using a much simplified photochemical scheme⁶. Namely, maximum depletions are found in high latitudes where the total ozone is largest; smallest depletions occur in equatorial latitudes where ozone is a minimum throughout the year. Figure 2 shows two calculations, using the depletions found in Fig. 1, of the

increase in solar UV radiation at 310 nm at the Earth's surface consequent on the reductions in ozone. The dashed line was obtained using the simple direct beam absorption scheme employed in the two-dimensional model, while the solid line uses the radiation scheme of Green *et al.*⁸, which includes the effects of scattering by air molecules and aerosols. It can be concluded immediately from Figs 1 and 2 that the ratio of ozone reduction to UV increase is not constant; the ratio varies with latitude and time. A value of two is perhaps appropriate to mid-latitudes during winter but it is clear that this value cannot be used generally.

It is desirable to know the possible increase in radiation at all wavelengths in the UV-B and ideally these increases should be weighted by the action spectrum for skin cancer. However, as this spectrum is not available we have used that of mild sunburn or erythema⁹. Ideally, we would wish to examine the

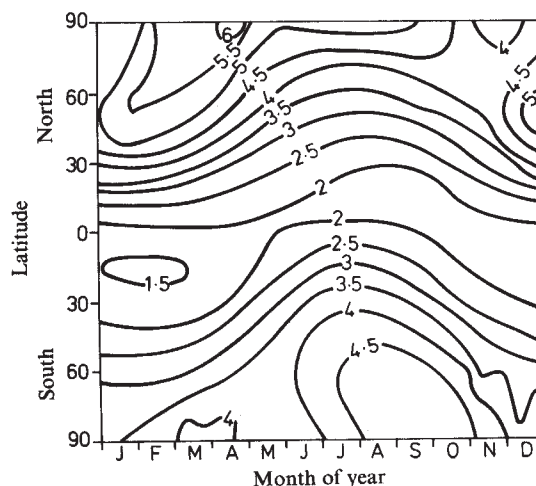


Fig. 1 Latitude-time cross-section of the percentage depletion in the total ozone amount in 1992 using the mean 1973–76 CFC production rates and the Oxford University two-dimensional model.

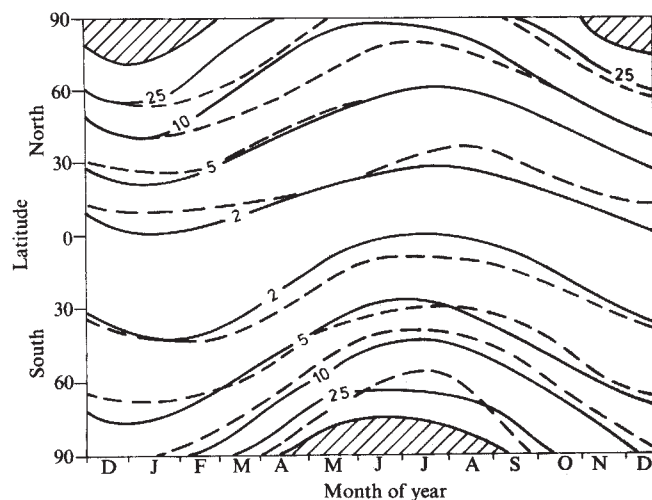


Fig. 2 Percentage change in UV radiation intensity in a narrow wavelength range centred at 310 nm resulting from CFC usage at 1992. Dashed line, calculations using a two-dimensional model; solid line, calculations using a simple UV radiation scheme.

perturbation in erythemally-weighted UV-B dose when the ozone depletion from CFC release reaches its steady state. Because of the many tens to hundreds of years that would be required and because of the detailed horizontal and vertical coverage in the two-dimensional calculations, the cost of running the model to steady state was considered prohibitive. However, the extrapolation of such predictions can be readily achieved by simple scaling up Fig. 1. A steady-state global ozone depletion of $13 \pm 3\%$ was selected on the basis of several UK modelling studies¹⁰. The slight annual increase in ozone depletion apparent in Fig. 1 during 1992, amounts to about a

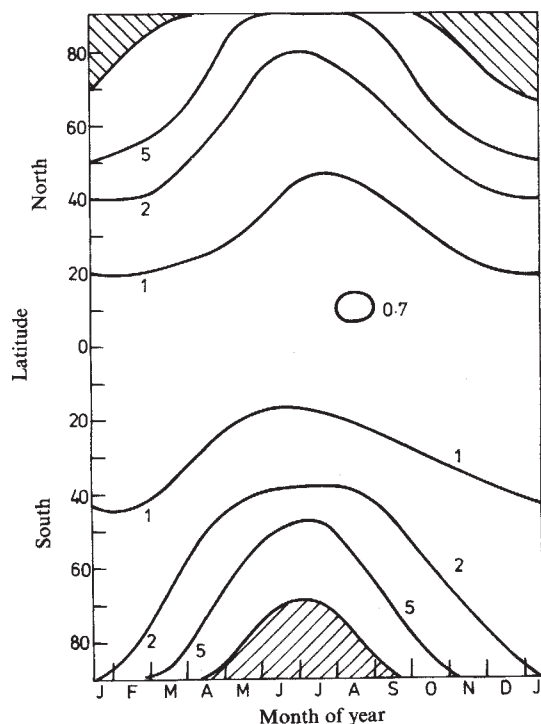


Fig. 3 Ratio of increase in erythemally-weighted UV-B dose to decrease in total ozone amount resulting from CFC usage.

factor of 1.1–1.2. This slight upward drift when extrapolated into steady state introduces errors in the ozone depletion which are small in comparison with the errors in the chosen value itself.

The results shown in Fig. 1 were extrapolated to steady state with respect to CFC release, giving a global ozone depletion of $\sim 13\%$. The radiation calculations including scattering were then used to compute the increase in the erythemally-weighted UV-B dose at steady state. A global increase of about 12% was found. Figure 3 shows the latitude-time section of the ratio of the percentage increase in UV-B dose to ozone depletion. Only in middle latitudes during winter does the ratio approach a value of 2 and for much of the globe the value is < 1 , such that the global ratio, as implied above, is about unity.

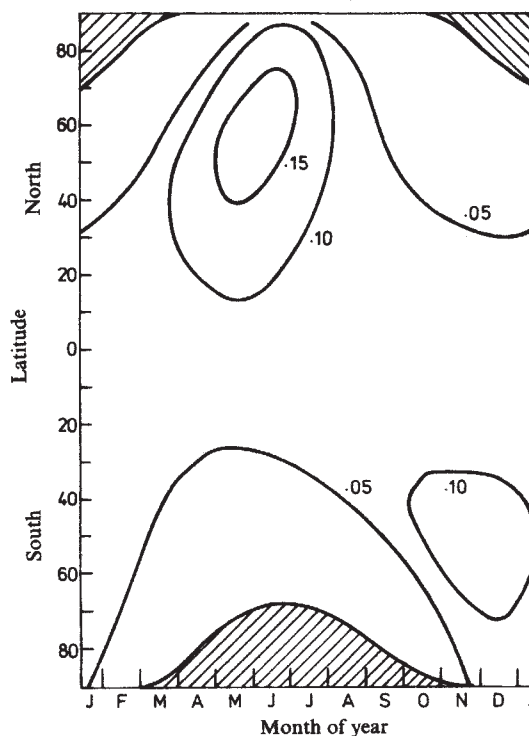


Fig. 4 Ratio of increase in erythemally-weighted UV-B dose from CFC usage to the dose at the Equator.

In high latitudes, the ratio does become larger and the significance of this is interesting. In Fig. 4 the increase in erythemally-weighted UV-B dose at the extrapolated steady state is compared with the dose at the Equator. The larger percentage increases in high latitudes seem to be of little significance when compared with the radiation reaching equatorial latitudes. However, plant and animal life in these high latitude regions coexist with small radiation doses so that we cannot necessarily infer that these changes will have only a small effect.

We conclude that no simple relationship exists between decreased stratospheric ozone and increased surface UV-B dose. The results of radiation calculations in steady-state one-dimensional models for a particular latitude should be treated with caution. Significant variations in the increased dose as a function of latitude and season have been demonstrated here. Inclusions of latitudinal and temporal effects will evidently be important in assessing the biological and botanical effects of possible future ozone depletions.

This work was sponsored by the Department of Environment.

Received 22 February; accepted 25 April 1980.

1. Johnston, H. S. *Science* **173**, 517 (1971).
2. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
3. Crutzen, P. J. *Ambio* **3**, 201–210 (1974).
4. Council on Environmental Quality, Fluorocarbons and the Environment (USGPO, Washington DC, 1975).
5. Hammon, A. L. & Maugh, T. H. *Science* **186**, 335–338 (1974).
6. Pyle, J. A. *Nature* **271**, 42–43 (1978).
7. Pyle, J. A. *WMO Publ. No. 511*, 239–246 (1978).
8. Green, A. E. S., Sawada, T. & Shettle, E. P. *Photochem. Photobiol.* **19**, 251–259 (1974).
9. Komhyr, W. D. & Machta, L. *The Perturbed Troposphere of 1990 and 2020* (Department of Transportation, Washington DC, 1973).
10. Department of Environment, *Chlorofluorocarbons and their Effect on Stratospheric Ozone* (HMSO, London, 1979).

ELF radio signals in the auroral ionosphere generated by non-linear demodulation of LF and/or MF transmissions

T. Turunen

Geophysical Observatory, Sodankylä, Finland

P. S. Cannon* & M. J. Rycroft†

Department of Physics, The University, Southampton, UK

We have conducted experiments measuring ELF and VLF radio signals in the afternoon to early morning (local time) between 24 March and 4 April 1979 in Northern Scandinavia. Apart from those of a natural origin (atmospherics, chorus, and so on), we received timing signals, six pips which occurred on the hour UT; the duration of each pip, of frequency $1\text{ kHz} \pm 0.5\text{ Hz}$, was $105 \pm 8\text{ ms}$. A spectrogram of these is shown in Fig. 1. We have established that these are not due to non-linear demodulation of strong higher frequency radio signals in any part of the equipment. We consider that they are produced by non-linear demodulation of signals from two or more amplitude modulated transmitters in the USSR operating at 173, 200, 236, 263 and 657 kHz. The demodulation occurs only during enhanced auroral electrojet activity.

The equipment primarily used in this study involved three goniometer ELF/VLF receivers, one located at Sodankylä (Finland), one at Tromsø (Norway) and one at Arjeplog (Sweden). All stations are within the auroral zone. The goniometer system, described in detail by Bullough and Sagredo¹ and Jarvis², uses loop antennas, of area 58 m^2 ; the goniometer modulation frequency was 27.5 Hz . Some measurements were made in parallel at Sodankylä using a VLF receiver designed at the Sodankylä Geophysical Observatory. This receiver was fed from a long (500 m) wire antenna. Additionally an ELF receiver (bandwidth $1\text{ Hz} - 1\text{ kHz}$), fed by a multiturn loop antenna, was operated during certain periods at Tromsø. Pips were detected by all these independent measuring systems. The ELF/VLF goniometer at Arjeplog was unfortunately disturbed by local interference at 1 kHz ; thus recordings made there could not be used for detailed study.

As well as the timing pips, distorted music, distorted speech and station identifiers were occasionally heard. Broadcast band monitoring and subsequent reference to the World Radio and TV Handbook³ identified the Russian First and Second Programmes on 173, 200, 236, 263 and 657 kHz as being the cause of these signals and, presumably, of the pips; timing signals were

being transmitted on at least some of the above mentioned frequencies every time the 1-kHz pips were detected. Some of the stations transmitting these programmes are within 1,500 km of Sodankylä; these are Kaliningrad (173 kHz, 1 MW), Moscow (173 kHz, 500 kW), Syktyvkar (173 kHz, 300 kW), Leningrad (200 kHz, 150 kW), Moscow (200 kHz, 100 kW), Leningrad (236 kHz, 1 MW), Arkhangelsk (236 kHz, 150 kW), Moscow (263 kHz, 2 MW) and Murmansk (657 kHz, 150 kW) (Fig. 2). The field strength of the 173-kHz transmissions was measured by J. Sivusaami (personal communication) at Sodankylä during the daytime on 3 April 1979, and found to be $112\text{ }\mu\text{V m}^{-1}$.

After completing the field experiments, 1 kHz amplitude modulated signals were injected into the goniometer receiving system, using a dummy aerial. Even with signals 52 dB stronger than those measured at 173 kHz, no demodulated output at 1 kHz was noted. Thus we conclude that the receiving system was linear and that the demodulation observed did not take place in the equipment.

The period of the experiment was characterized by two different levels of local geomagnetic activity. From 14 to 21 March, when no pips were recorded, the daily sum of K_{HDZ} indices, measured at the Sodankylä Geophysical Observatory, varied between 2 and 20. From 22 March to 4 April the corresponding values varied between 20 and 45 and, with the exception of 1 April, pips were recorded at least once each day. We conclude that pips were recorded only when the daily magnetic activity was moderately high. The probability of pips occurring is greatest in the local evening sector, at 19 and 20 UT. Only one measurement was available at 18 UT. Inspection of STARE (Scandinavian Twin Auroral Radar Experiment⁴) data and magnetograms from Sodankylä and Kiruna shows that an active auroral zone current system was always nearby during the times of pip occurrence. Additionally, in cases when the source location was possible, this location probably fell within the eastward electrojet.

On six occasions the signal-to-noise ratio was high enough to allow the determination of the azimuthal bearings of the 1-kHz signals both from Sodankylä and Tromsø, enabling the source to be located by goniometer triangulation. On five occasions the sources were approximately the same. Figure 3 shows this location with the worst case envelope of standard errors. The sixth location is also shown with its error diamond. On three more occasions a bearing angle, with 180° ambiguity, was obtained from Sodankylä. These directions are also shown in Fig. 3. They do not coincide with the direction of the five cases having roughly the same source location. The group of five events from the same location occurred at 19 or 20 UT; the sixth location was determined at 22 UT. Two of the directions determined from Sodankylä only are at 19 UT and the other at 21 UT. Two of the measurements belonging to the group of five events of the same source location were measured on the same day.

We summarize the results obtained as follows. During times of active auroral zone currents, and providing that the daily magnetic activity has been relatively high, there is a high probability of the occurrence of pips in Northern Scandinavia. These signals are considered to be due to a demodulation (self-detection) effect (in the lower ionosphere) of certain LF (and MF) AM broadcasting stations transmitting hourly time marks at 1 kHz within the programme schedule of the Russian Internal Service First and/or Second Programmes. The probability of their occurrence is highest around 19 and 20 UT, but a high probability at 18 UT is not ruled out, because of a data gap. At these times the signal-to-noise ratio is sometimes good enough to indicate that the signal source has a favoured location in the ionosphere (shown in Fig. 3) and that this location does not change with time; there are also other possible locations of the source.

Experimental evidence for the demodulation of HF radio signals in the presence of auroral electrojets has been reported by Kapustin *et al.*⁵. Theoretical treatments for the generation of ELF/VLF waves in the auroral ionosphere have been published

Present address: *Marconi Space and Defence Systems, Portsmouth, UK; †NERC British Antarctic Survey, Cambridge, UK.

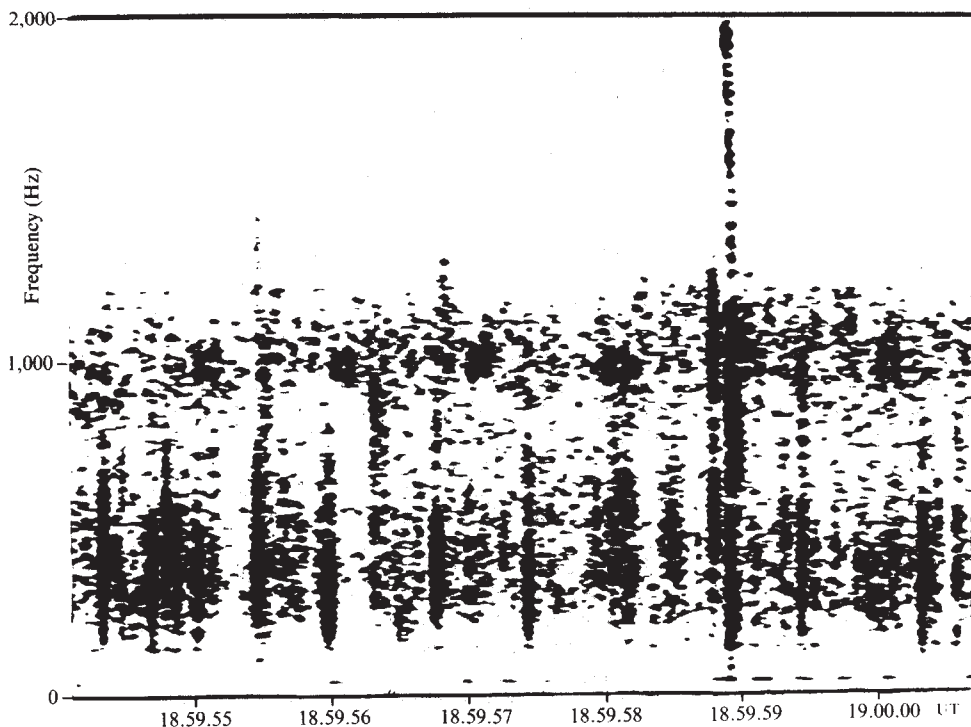


Fig. 1 Spectrogram showing atmospheric and six 1-kHz pips; frequency splitting of the signal, into upper and lower sidebands 55 Hz apart, is due to the goniometer modulation.

by Stubbe and Kopka^{6,7} and by Kapustin *et al.*⁵, and in the references cited therein. It is proposed there that ELF/VLF radio waves can be generated after heating the lower ionosphere with a strong HF wave having amplitude modulation at ELF/VLF. The modulation causes periodic changes in electron temperature which cause corresponding changes in the Hall and Pedersen conductivities. As a consequence, an a.c. component, at the frequency of the periodic heating, is formed in the auroral electrojet, which radiates a wave at the heating frequency. Stubbe and Kopka^{6,7} consider the case when a limited area of auroral ionosphere is heated by a vertical narrow beam radio transmission at HF which can penetrate both the D and E layers in most conditions. This also applies to the experimental arrangements of Kapustin *et al.*⁵.

We believe that we have found a basically similar phenomenon, but caused by obliquely propagating amplitude modulated LF and/or MF transmissions acting in the lower portion of the auroral D-layer. In the case of one transmitting

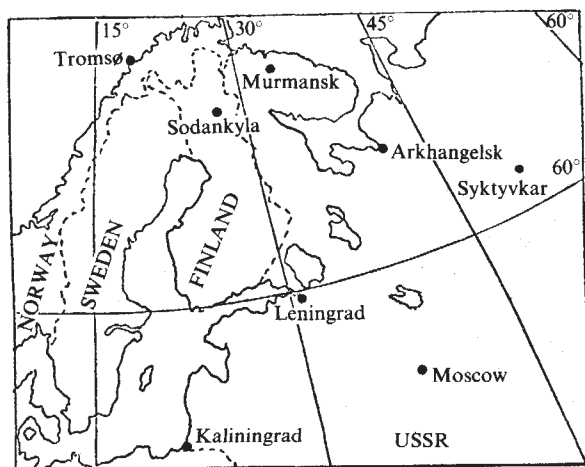


Fig. 2 Mercator projection of Scandinavia and North Western USSR, showing Soviet high power LF and MF broadcast stations within 1,500 km of Sodankyla, on 173, 200, 236, 263 and 657 kHz.

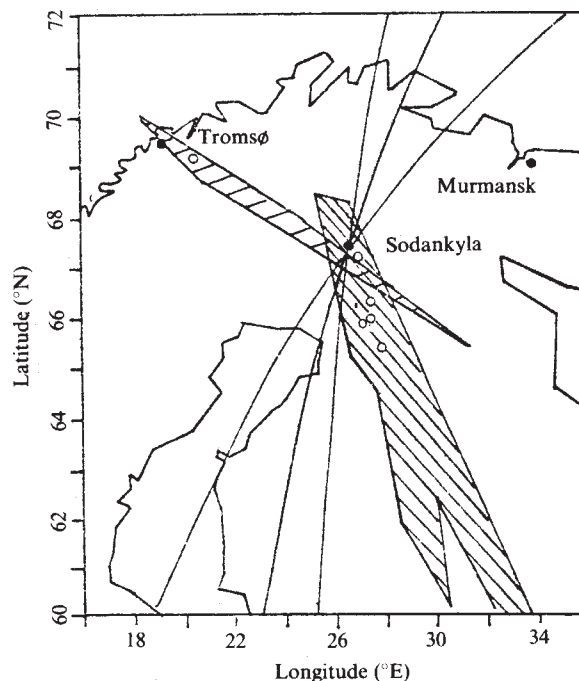


Fig. 3 Cartesian projection of Northern Scandinavia showing the positions (O) of the sources in the ionosphere of the 1-kHz pips, with the worst case error envelopes (hatched area). Additionally, three bearings with 180° ambiguity and a typical error of $\pm 6^\circ$, are shown from Sodankyla.

station, maxima of the heating effect are located at any instant at positions where the maxima of the modulation of the direct wave and waves after experiencing several sky and ground reflections are essentially in phase with each other. When the same programme is transmitted from two stations an interference pattern is formed, with maxima in the heating power moving along roughly hyperbolic trajectories in space. The positions of these trajectories depend on the modulation frequency. Non-linear demodulation occurs only in regions where the appropriate ionospheric conditions are satisfied.

In the case of three stations there are a few particular areas, which remain fixed, where the crests of the modulated waveforms occur and where the periodic heating maximizes. For four or more stations, the extra stations may, depending on their locations and the phases of modulation of their transmitted signals, contribute to the phenomenon.

We consider that at 1 kHz a proper phasing of the signals from at least three stations is obtained in the area in which we have located five sources. We then explain our discovery that pips occur only when the daily magnetic activity is above a certain threshold by this preferred area lying in the electrojet. On quiet days the main current system lies at higher latitudes beyond the particular area mentioned. The distortion of music or speech occurs because sources at different modulating frequencies have different fixed locations; not all of these may lie in areas suitable for ELF/VLF wave generation. This gives rise to a complicated and spectrally varying weighting of the signal. Other such areas may also be invoked to explain the sixth location shown in Fig. 3 and the three signals, observed only at Sodankyla, which had quite different propagation directions.

We thank our colleagues who took part in the present observational programme, and also members of the Measuring Station of the Finnish Broadcasting Company for measurements of various LF and MF transmitter field strengths.

Received 28 January; accepted 5 June 1980.

1. Bullough, K. & Sagredo, J. L. *Planet. Space Sci.* **21**, 899 (1973).
2. Jarvis, M. J. thesis, Univ. Southampton (1976).
3. *World Radio and TV Handbook* (Billboard, London, 1979).
4. Greenwald, R. A., Weiss, N., Nielsen, E. *Radio Sci.* **13**, 1021 (1978).
5. Kapustin, I. N. et al. *JETP Lett.* **25**, 228 (1977).
6. Stubbe, P. & Kopka, H. *J. geophys. Res.* **82**, 2319 (1977).
7. Stubbe, P. & Kopka, H. *Ionospheric Modification Experiments in Northern Scandinavia—a Description of the Heating Project* (Max Planck Institute für Aeronomie, MPAE-W-02-79-04, 1979).

The pear-shaped section of the Earth

D. G. King-Hele*, C. J. Brookes† & G. E. Cook*

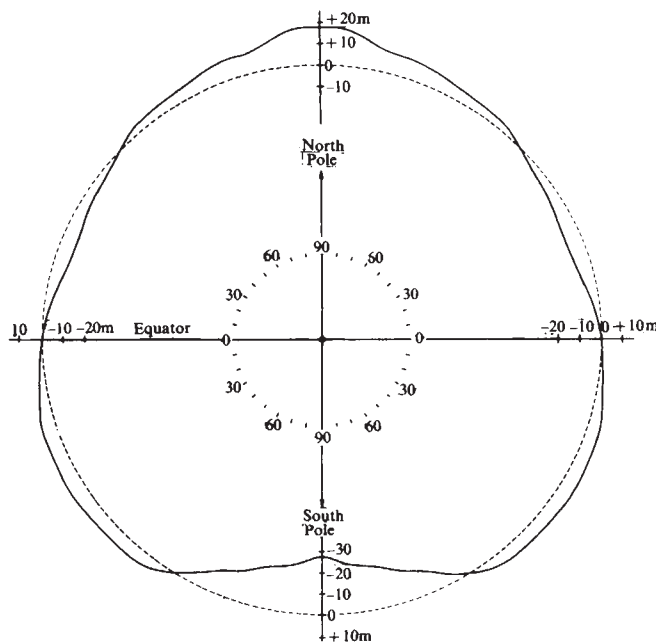
* Royal Aircraft Establishment, Farnborough, Hants GU14 6TD, UK
† University of Aston, Gosta Green, Birmingham B4 7ET, UK

The Earth's gravitational field, and hence the shape of the Earth's sea-level surface, the geoid, can be determined from analysis of the motion of artificial satellites in orbit about the Earth. The method is particularly powerful in revealing the longitude-averaged meridional profile of the geoid, that is, an average slice through the poles, which from its north-south asymmetry is often called 'pear-shaped'. We now report an improved determination of this profile, which should be accurate to about 50 cm at latitudes up to 86°, and replaces our previous evaluation^{1,2}.

The Earth's gravitational potential may be expressed as the sum of (1) an infinite series of zonal harmonics independent of longitude, and (2) a double infinite series of tesseral harmonics, dependent on both latitude and longitude. Here we are only concerned with (1), and we express the longitude-averaged potential at an exterior point distant r from the Earth's centre at geocentric latitude ϕ in the standard form³

$$\bar{U} = \frac{\mu}{r} \left\{ 1 - \sum_{l=2}^{\infty} J_l \left(\frac{R}{r} \right)^l P_l(\sin \phi) \right\} \quad (1)$$

where μ is the gravitational constant for the Earth, $398,600 \text{ km}^3 \text{ s}^{-2}$, R is the equatorial radius, 6,378.14 km, and $P_l(\sin \phi)$ is the Legendre polynomial of degree l and argument $\sin \phi$. The J_l are constant coefficients: $J_2, J_4, J_6, \dots$ express the effects of harmonics symmetrical about the Equator; $J_3, J_5, J_7, \dots$ express the effects of north-south asymmetry, and it is



the south polar radius, is 45.1 m, compared with 44.7 m from the previous solution.

We have also derived 8- and 14-coefficient solutions as alternatives, but they give almost identical results, with differences of less than 30 cm in geoid height up to 86° latitude.

Comparison with the corresponding geoid profile from the most recent comprehensive gravity field⁷, GEM 10B, which has a completely independent set of odd zonal harmonics, shows differences of less than 80 cm up to latitudes of 85°, the average difference being about 25 cm. The effect of changing the even harmonics from those of GEM 10B to the most recent set derived at the Smithsonian Astrophysical Observatory shows an average difference of only 20 cm. For these and other reasons, we suggest that the profile of Fig. 1 should be accurate to 50 cm up to latitudes of 86°. Over the remaining 0.2% of the Earth's surface very near the poles, the errors may be larger, possibly up to 1.5 m: basically, this increased error arises because there is so little of the Earth's mass at these latitudes.

Although the longitude-averaged geoid profiles agree well, the predicted amplitudes of the oscillations in perigee distance do not: at certain inclinations, the amplitudes given by our zonal harmonics differ from those given by GEM 10B by up to 350 m, even if the singularity near 63.4° inclination is excluded. So further improvements in the values of the zonal harmonics, utilizing satellites in chosen orbits specially observed, will be needed if orbital models capable of centimetric accuracy, to match centimetric laser observations, are ever to be developed.

Received 28 April; accepted 3 June 1980.

1. King-Hele, D. G. & Cook, G. E. *Nature* **246**, 86 (1973).
2. King-Hele, D. G. & Cook, G. E. *Planet. Space Sci.* **22**, 645 (1974).
3. Merson, R. H. & King-Hele, D. G. *Nature* **182**, 640 (1958).
4. Brookes, C. J. & Holland, D. *Planet. Space Sci.* **26**, 611 (1978).
5. Brookes, C. J. *Planet. Space Sci.* **24**, 711 (1976).
6. King-Hele, D. G., Brookes, C. J. & Cook, G. E. *Geophys. J.R. astr. Soc.* (in the press).
7. Lerch, F. J. et al. *Paper Spring A. Meet. Am. Geophys. Un. Florida* (1978).

West Antarctic ice sheet fluctuations in the Antarctic Peninsula area

D. E. Sugden & C. M. Clapperton

Department of Geography, University of Aberdeen,
Aberdeen AB9 2UF, UK

The West Antarctic ice sheet is believed to be inherently unstable because much of it is grounded below sea level¹⁻⁴. It has been suggested that the ice sheet has withdrawn from its late Wisconsin maximum position, grounded at the edge of the continental shelf, and is now undergoing collapse as a delayed response to the warming and sea-level rise of the Holocene⁵, and that the ice sheet is likely to collapse shortly in response to rising CO₂ levels in the atmosphere⁶. We present here some geomorphological evidence from Alexander Island and the Antarctic Peninsula which does not agree with either hypothesis. Rather, following deglaciation from the Wisconsin maximum, there was less ice than at present around 8,000 yr ago. The ice shelf in George VI Sound has built up subsequently.

Alexander Island is a mountainous block of country approximately 450 km from north to south and 300 km from east to west. Jagged mountains rise to altitudes exceeding 3,000 m in the north of the island but more subdued tracts of dissected plateau 250–800 m altitude are more characteristic of the south and east. These topographical contrasts partly reflect the island's geology for, whereas the spectacular Elgar Mountains of the north are composed of plutonic rocks, the bulk of the island's plateau scenery in the south and east coincides with extensive

areas of sedimentary strata, particularly shales, sandstones, volcanoclastics and conglomerates⁷⁻⁹. The adjacent part of the Antarctic Peninsula, similarly aligned from north to south, is 200 km across and includes mountains rising to 3,000 m. As on Alexander Island, the land is more extensively buried by ice fields towards the south. The Peninsula is geologically distinct from Alexander Island in that the bulk of it is composed of plutonic and volcanic rocks and a variety of gneisses, granites and lavas predominate¹⁰. It is believed that block faulting has separated the two land areas resulting in the trench of George VI Sound which is over 20 km wide, and 500 km long.

The Peninsula is covered by a northward extension of one of the three main ice domes comprising the West Antarctic ice sheet and from its axis the ice drains westwards and eastwards into ice shelves. The western ice shelf, 100–500 m thick, flows westwards across George VI Sound and abuts the east coast of Alexander Island where it is depositing a conspicuous moraine. This ice shelf terminates at either end of the Sound in environmental conditions where summer temperatures rise to within a few degrees of 0 °C. As such the ice shelf in George VI Sound is in a marginal environment for its survival¹¹. Alexander Island itself supports an ice cap and ice fields which discharge outwards to each coast. Outlet glaciers flowing into George VI Sound ice shelf are quickly dominated by the more powerful ice flow from the Peninsula and are almost immediately deflected northwards and southwards.

Evidence of past glacial maximum conditions points to increased ice thicknesses on both Alexander Island and the Antarctic Peninsula, with both areas acting as independent centres of outflow. On eastern Alexander Island the evidence is in the form of glacial trough alignment, the direction and position of striations on plateau surfaces and interfluvies, and erratic fans which can be traced from distinctive bedrock outcrops. Examples of these features in the Ablation Point area show how flow was consistently eastwards towards George VI Sound (Fig. 2). Compatible with this reconstruction is the absence of Peninsula erratics on the eastern plateaux of Alexander Island; although Peninsula-derived granites and gneisses are clearly identifiable in the moraine flanking the present ice shelf in George VI Sound, none were found above an altitude of 85 m. On Plateau surfaces in the Batterbee Mountains of the Peninsula across from Alexander Island the flow of a former ice cover was westwards into the Sound and then northwards parallel with the Sound. Clear evidence for this comes from roches moutonnées, striations, sub-glacial meltwater channels and the slope of high-level lateral moraines which cross the plateau at 400–670 m altitude (Fig. 3).

Previous reconstructions of maximum ice conditions in the late Wisconsin for this part of Antarctica show the ice flowing from the Peninsula axis across Alexander Island to the edge of the continental shelf on its west side^{5,12}. The field evidence from eastern Alexander Island indicates ice-flow directions directly opposite to those implied by these previous reconstructions and implies that the ice masses were thinner, consisting of separate domes that were based on Alexander Island and on the Peninsula and becoming confluent in George VI Sound.

Some indication of the age of the last maximum glaciation was obtained from a shelly till deposit found on a spur 3 km north of Two Step Cliffs and lying between 94 and 114 m above the present ice shelf. The shells consist of fragments of *Hiatella solida*. They lie in a till with striated clasts and a typical glacial matrix (mean ϕ 0.21, sorting 4.05, skewness 0.39, kurtosis 0.82). Analysis of the lithology of the clasts and of the magnetic susceptibility of the matrix suggests the till is derived from sedimentary rocks characteristic of Alexander Island and is quite different to the Peninsula-derived ice shelf moraines in the vicinity originating on igneous rocks. The outer fraction of the shells gave a radiocarbon age of $32,160 \pm 360$ yr and the inner fraction $30,600 \pm 660$ yr (SSR 1499). That such an old radiocarbon age is likely to be a minimum was confirmed by the results of amino acid analysis on three separate shells which yielded total D/L ratios of 0.169, 0.113 and 0.147 (AAL 1297).

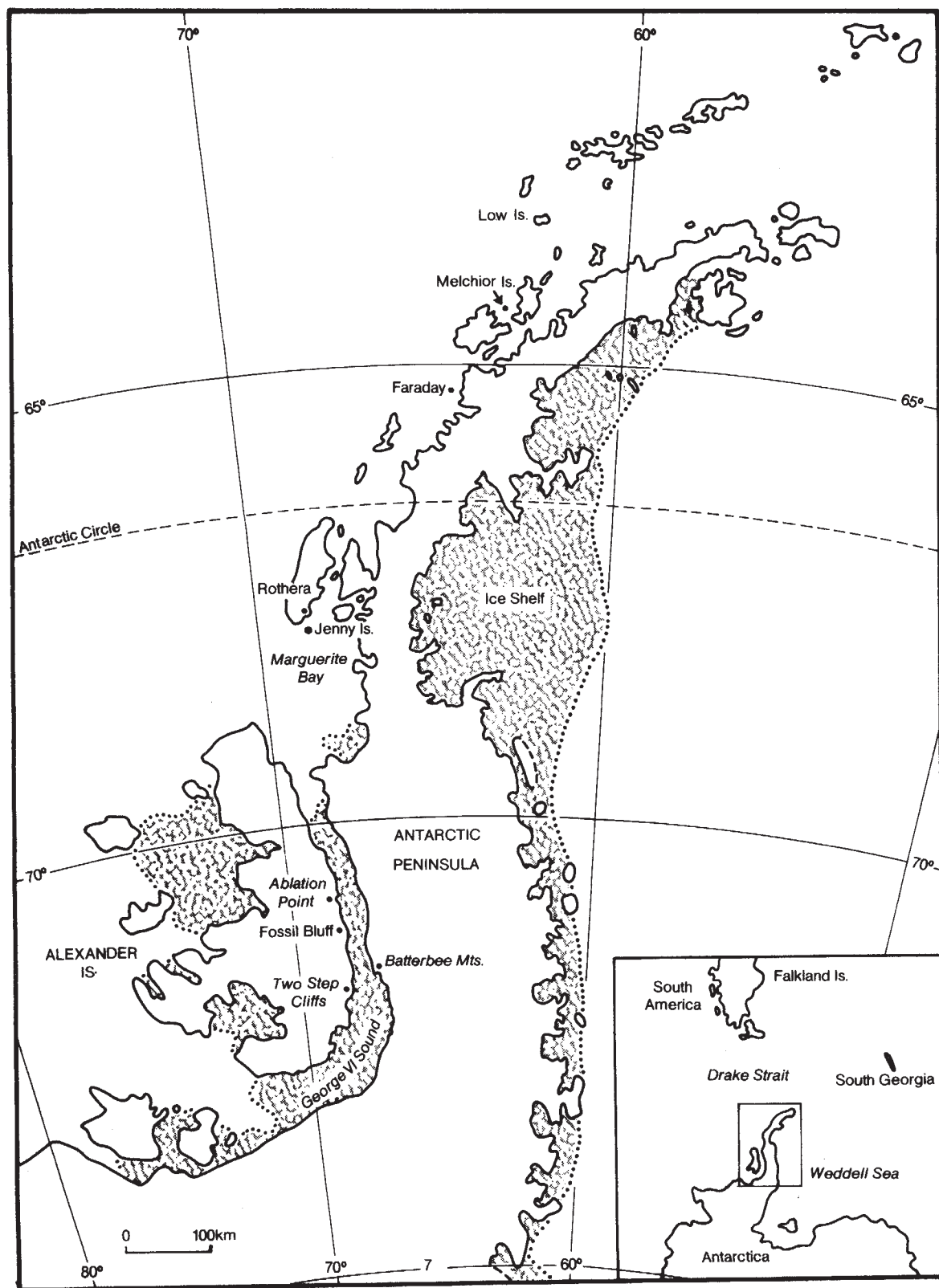


Fig. 1 The Alexander Island–Antarctic Peninsula area of West Antarctica.

We cannot place accurate absolute dates on these values without knowing the precise thermal history of the shells since their death. Nevertheless, a reconstruction of the most likely thermal history (in seawater or beneath a warm-based glacier) produces an age which is compatible with that of the last interglacial (G. H. Miller, personal communication). We can go even further and suggest that at the time of their growth the Sound was free of an ice shelf. The basis of this suggestion is the discovery from the J9 drill hole through 420 m of the Ross Ice Shelf, 430 km from the open sea, that no living infauna existed and that the sediments contained no animal traces or bioturbations¹³. As George VI Sound Ice Shelf is 300 m thick over 200 km distant from the open sea in the vicinity of Two Step Cliffs, it is reasonable to suggest that shells may be unable to thrive beneath its canopy and must, therefore, reflect open conditions in the Sound.

The most likely explanation of the shelly till is that it was formed by an enlarged ice sheet grounded in George VI Sound and just impinging on Alexander Island on the promontory north of Two Step Cliffs. In such a case the dates imply that the ice cover was substantially thicker in the Wisconsin than at present. In the absence of any alternative moraine limits it seems reasonable to attribute the maximum reconstruction described earlier to this Wisconsin advance.

Evidence of the Holocene history of the ice shelf comes from the moraine which flanks the ice shelf where it abuts onto Alexander Island. The moraine has two components. One is a landward ridge 5–30 m high which apparently contains no ice. The other is immediately adjacent and forms one to three ice-cored ridges. The crests of all the ridges are lower than the overall surface of the ice shelf. At Two Step Cliffs barnacles *Bathylasma corolliforme* (Hoek) were found in the landward of

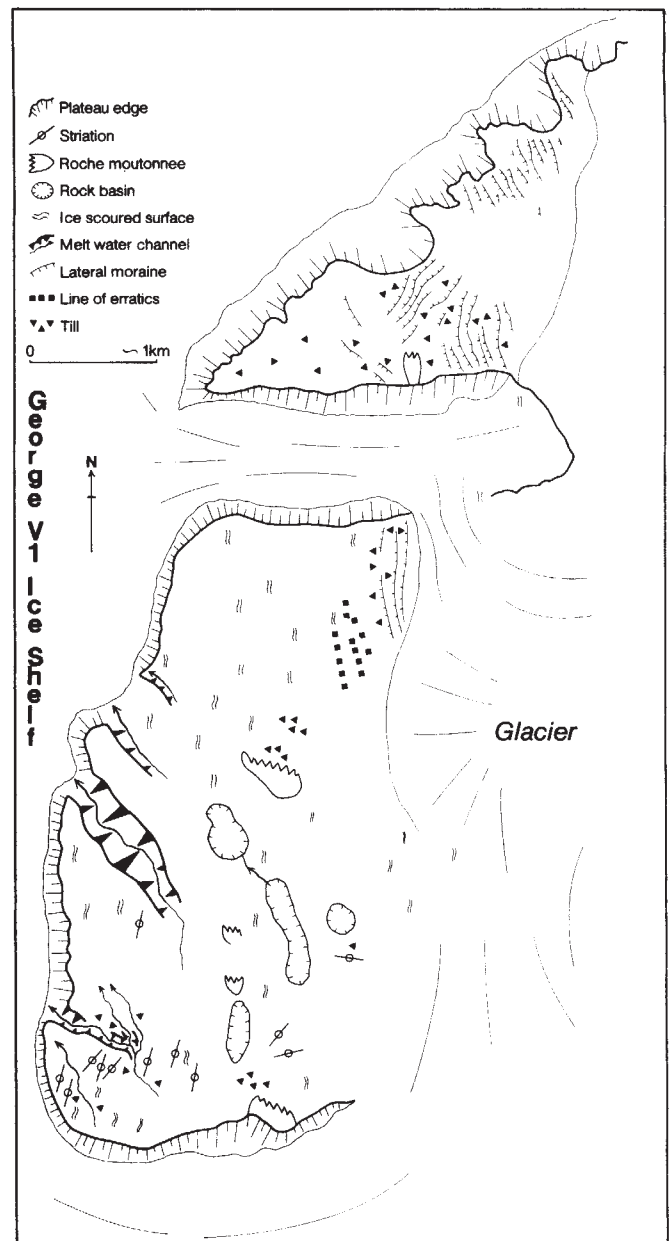


Fig. 3 Evidence of former ice movements sub-parallel to George VI Sound in the Batterbee Mountains area.

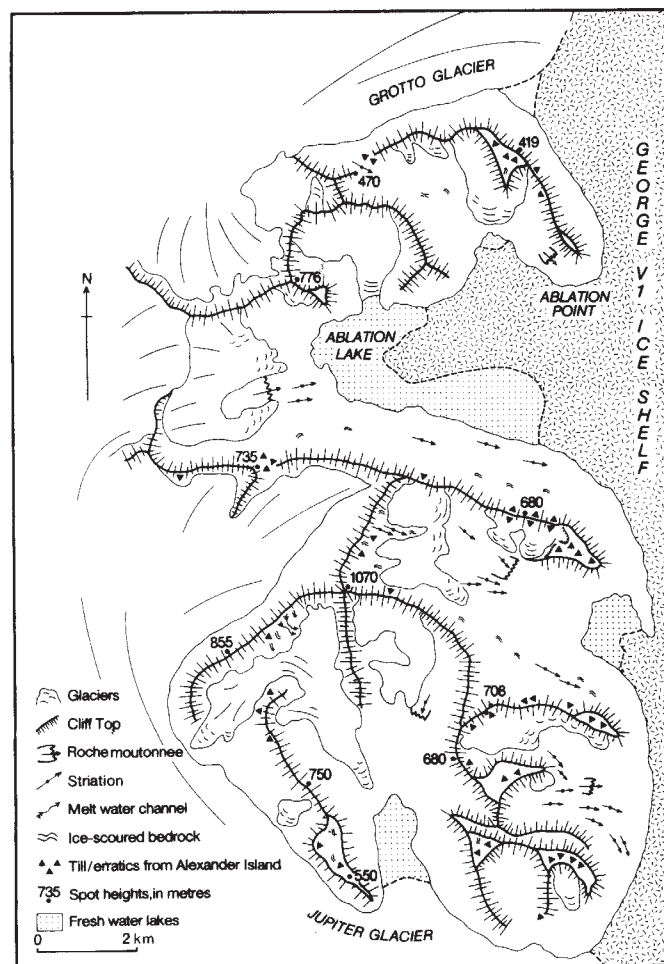


Fig. 2 Evidence of former ice movement from west to east across the Ablation Point area.

the two ridges. The presence of striated clasts and mechanical analysis of the matrix confirm that the deposit is till but with signs of water activity (mean ϕ 2.09, sorting 4.19, skewness -0.02 , kurtosis 0.96). Radiocarbon dating of the outer fraction of the barnacles produced an age of $6,930 \pm 60$ and the inner fraction $7,200 \pm 50$ (SRR-1500). If one allows for a correction factor of ~ 750 yr for the ^{14}C deficiency in Antarctic waters based on radiocarbon dates on modern shells in the Antarctic Peninsula area^{14,15} then the adjusted dates come out as 7,700 and 7,950 radiocarbon yr BP. This Holocene age is fully consistent with the results of amino acid analysis which revealed total D/L ratios of 0.028 and 0.01 (AAL 1298). The position and age of the barnacles implies that they were growing in George VI Sound around 8,000 yr ago and that subsequently they have been incorporated in the ice shelf moraine. There are two possible implications pertinent to the Holocene history of the ice shelf. The most likely is that the barnacles were growing in an ice-free Sound $\sim 8,000$ yr ago and the ice shelf has built up subsequently. The other less likely possibility is that the

barnacles were growing beneath the ice shelf; in this case the implication is that the ice shelf has not been materially thicker than at present for ~8,000 yr (although it is possible that it has been thinner and less extensive).

These results of investigations in the Alexander Island and Antarctic Peninsula area yield five conclusions relevant to the history of the West Antarctic ice sheet.

- (1) During the last maximum glaciation separate ice domes were centred over both Alexander Island and the Peninsula. This represents an ice thickness less than that predicted on theoretical grounds in the CLIMAP reconstructions¹².
- (2) Amino acid analyses on shells suggest that the last maximum was achieved in the Wisconsin and followed an interval when the Sound may have been free of ice, perhaps during the last interglacial.
- (3) George VI Ice Shelf, one of the most sensitive to environmental change in West Antarctica, has apparently reformed in the last 8,000 radiocarbon yr following a period of warmer conditions when the Sound was ice-free.
- (4) Ice sheet fluctuations in this part of West Antarctica, especially the signs of a warmer period ~8,000 yr ago, seem to be in phase with world environmental changes as reflected for example by deep sea cores¹⁶ and recent ice cores¹⁷.
- (5) With regard to the hypotheses concerning the possible instability of the West Antarctic ice sheet, one can conclude that in this sector: the ice in the last maximum was thinner than previously supposed and that there is no sign of progressive ice sheet collapse; rather there seems to have been an expansion in

the last 8,000 yr as George VI ice shelf reformed; the West Antarctic ice sheet is able to survive without the presence of George VI ice shelf; so long as the ice shelf survives, then there is no need to fear imminent West Antarctic ice sheet collapse due to climatic warming as proposed by Mercer.

We thank NERC for financial support, the British Antarctic Survey for logistic support, Drs G. Oliver and W. Newman for identifying the shells, Dr D. Harkness for radiocarbon dating, Dr J. T. Andrews for amino acid analysis and Professor F. Oldfield for magnetic susceptibility measurements.

Received 29 February; accepted 23 May 1980.

1. Hughes, T. J. *geophys. Res.* **78**, 7884–7910 (1973).
2. Weertman, J. *Nature* **260**, 284–286 (1976).
3. Hughes, T. *Rev. Geophys. Space Phys.* **15**, 1–46 (1977).
4. Thomas, R. H., Saunderson, T. J. O. & Rose, K. E. *Nature* **277**, 355–358 (1979).
5. Hughes, T. *Rev. Geophys. Space Phys.* **13**, 502–526 (1975).
6. Mercer, J. H. *Nature* **271**, 321–325 (1978).
7. Horne, R. R. *Br. Antarct. Surv. Bull.* **22**, 61–76 (1969).
8. Bell, C. M. *Br. Antarct. Surv. Bull.* **33** & **34**, 1–16 (1973).
9. Bell, C. M. *Br. Antarct. Surv. Bull.* **39**, 35–44 (1974).
10. Adie, R. J. *Antarctic Research* (eds Priestley, R., Adie, R. J. & Robin, G. de Q.) 118–162 (Butterworths, London, 1964).
11. Mercer, J. H. *Int. Ass. Sci. Hydr.* **79**, 217–225 (1968).
12. Denton, G. H. & Hughes, T. J. (eds) *The Last Great Ice Sheets* (Wiley-Interscience, in the press).
13. Lipps, J. H., Ronan, T. E. & DeLaca, T. E. *Science* **203**, 447–449 (1979).
14. Sugden, D. E. & John, B. S. *Palaeoecology of Africa and of the Surrounding Islands and Antarctica*, Vol. 8 (ed. van Zinderen Bakker, E. M.) 139–159 (Balkema, Cape Town, 1973).
15. Harkness, D. D. *Br. Antarct. Surv. Bull.* **47**, 43–59 (1979).
16. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
17. Loris, C., Merlivat, L., Jouzel, J. & Pourchet, M. *Nature* **280**, 644–648 (1979).

Shelly microfossils near the Precambrian–Cambrian boundary, Mackenzie Mountains, northwestern Canada

S. Conway Morris

Department of Earth Sciences, Open University, Milton Keynes MK7 6AA, UK

W. H. Fritz

Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8, Canada

Our understanding of early metazoan evolution is being revolutionized by recent discoveries of pre-trilobite faunas. A fascinating aspect of this evolution is the appearance and development of metazoan skeletal parts near the Precambrian–Cambrian boundary. Fossils of these parts are best known from the lowermost stage (Tommotian) of the Cambrian in Siberia which contain archaeocyathids and small shelly fossils, the latter commonly of phosphatic composition^{1–3}. Although the sudden, in terms of geological time, appearance of metazoan hard parts near the boundary is recognized as an outstanding evolutionary event, none of the proposals given to explain the ‘Cambrian explosion’ (see refs 4–6) has been widely accepted. Additional information from areas outside Siberia is now needed to broaden our knowledge on early Metazoa, and to assist a geological team, the Precambrian–Cambrian Boundary Working Group of the International Union of the Geological Sciences, that is locating, defining and correlating the Precambrian–Cambrian boundary on a world-wide basis. Here we report the discovery of a small fauna of microfossils near the boundary in the Mackenzie Mountains. New forms are present and a known species is used for correlation with the classic Precambrian–Cambrian sections in Siberia and successions elsewhere.

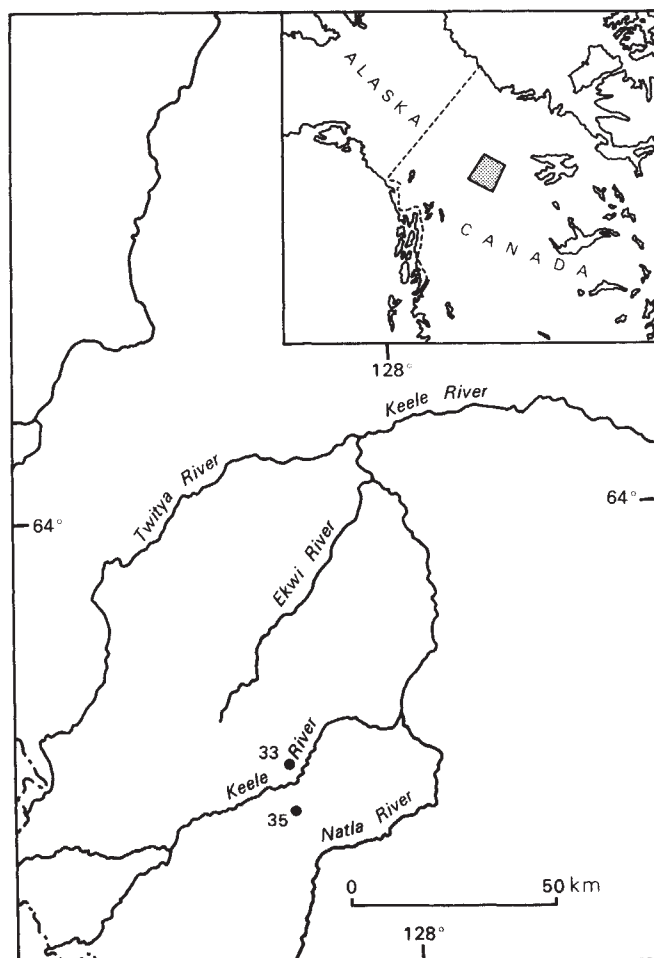


Fig. 1 Map locating sections 35, that contains shelly microfossils at GSC locality 94597, and 33. Section 35 is located between lat. 63°22'5', long. 128°38.75' (base of section) and lat. 63°22', long. 128°35.75' (top of section).

A sample from the Mackenzie Mountains (Geological Survey of Canada (GSC) locality 94597) collected by one of us (W.H.F.) yielded shelly microfossils after 660 g of limestone was digested in dilute acetic acid (~10%). The sample is from stratigraphic section 35 (ref. 7), located between the Keele and Natla Rivers, Northwest Territories (Fig. 1). Note that the sample came from a horizon 3,065 feet (934 m) stratigraphically below the next youngest fossil locality (GSC locality 94599) (Fig. 2) which is located in the Sekwi Formation^{7,8} and yields cf. *Parafallotaspis grata* Fritz, the oldest described trilobite species from North America. GSC locality 94597 is in the middle of a 135 foot (41.4 m) thick limestone unit that has been correlated with map

Unit 11 (ref. 7), an informal stratigraphic unit recognized by Blusson⁹ a short distance to the east. In addition to the microfossils documented here, stromatolites and burrows were noted⁷ at GSC locality 94597. These latter fossils, together with the light colour of the limestone, suggest the sampled limestone was deposited in shallow water.

The following microfossils were obtained: (1) A single specimen of *Protohertzina* cf. *P. anabarica* Missarzhevsky¹⁰ (Fig. 3a-c). Both ends are broken, but the specimen is otherwise well preserved. *P. anabarica* is the genotype and four other species have been described¹⁰⁻¹². The genus belongs to a group of phosphatic fossils known as protoconodonts that may form one

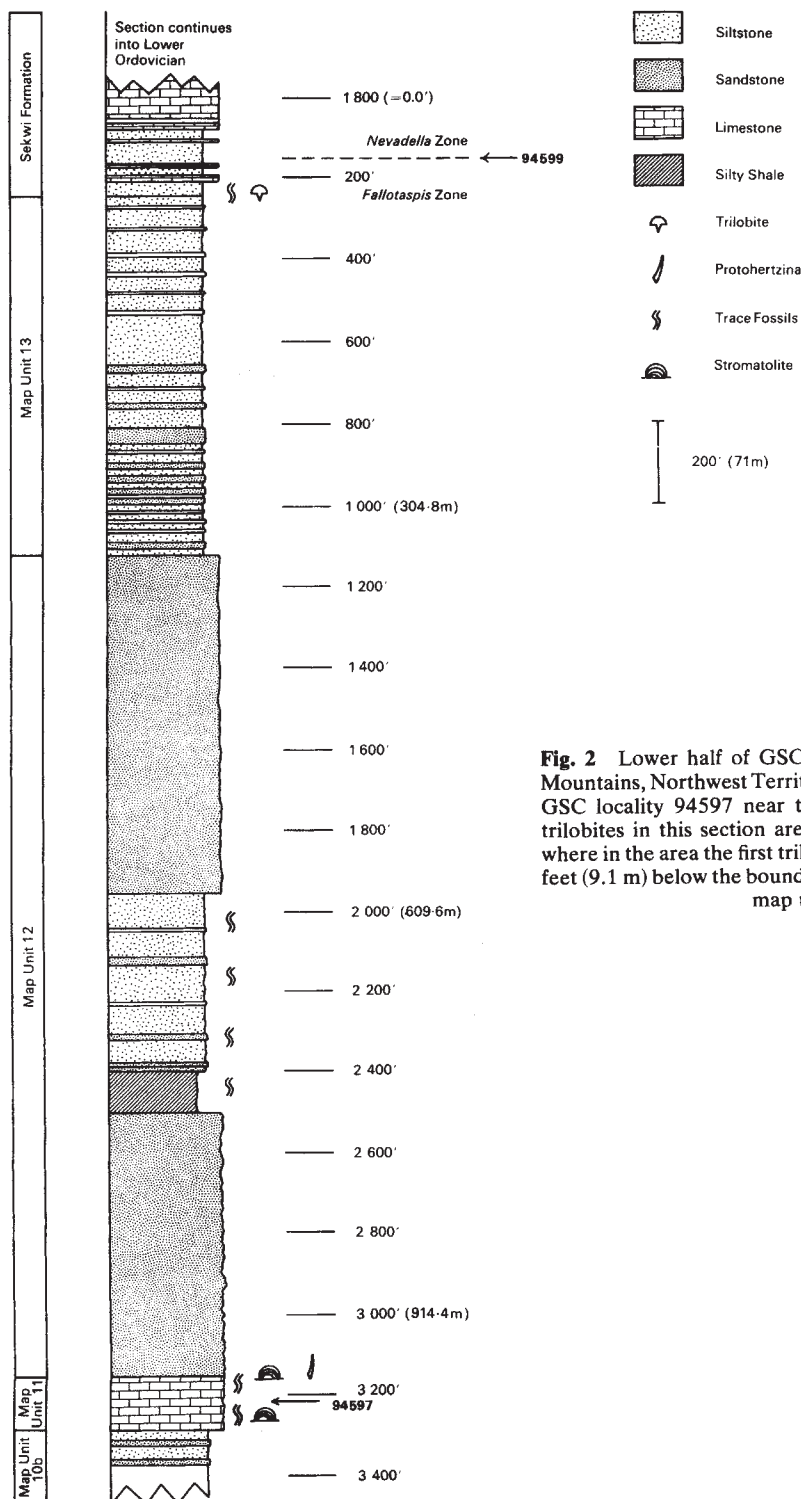
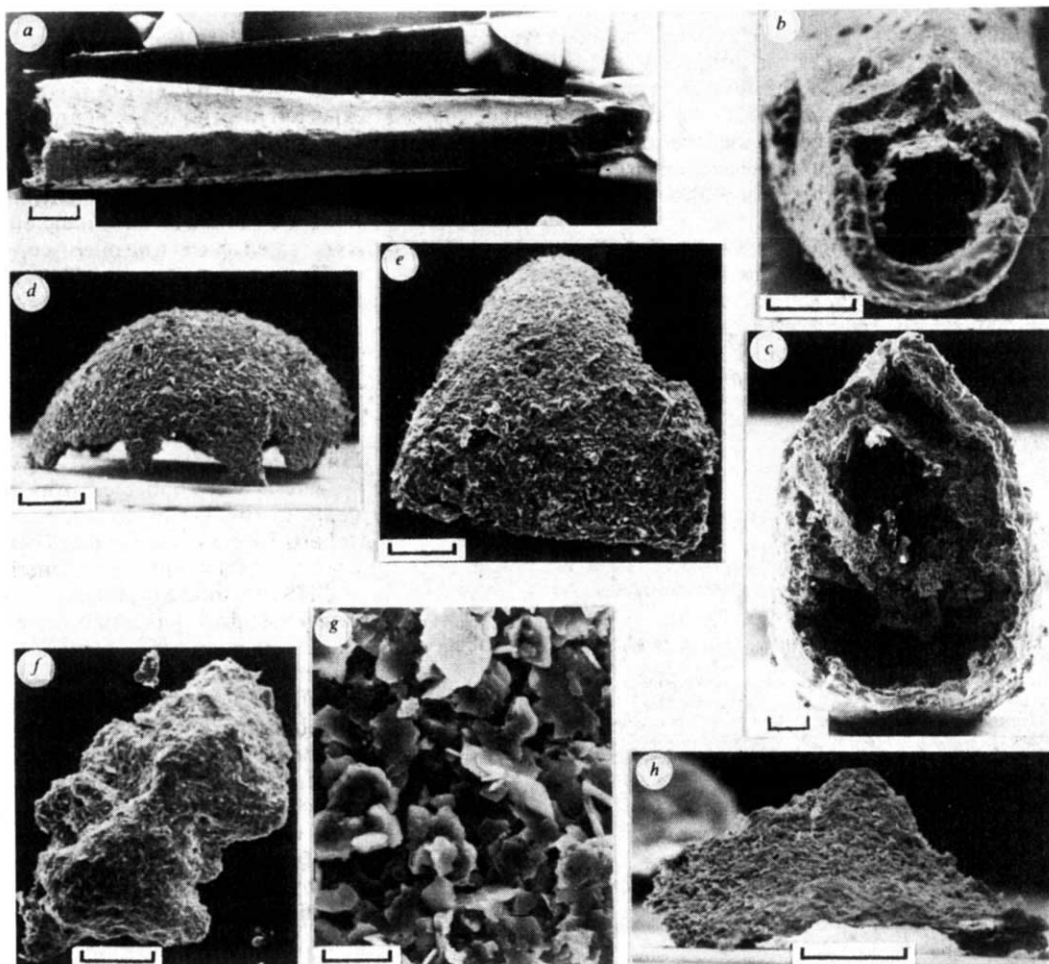


Fig. 2 Lower half of GSC stratigraphic section 35, Mackenzie Mountains, Northwest Territories. The shelly microfossils are from GSC locality 94597 near the base of the section. The earliest trilobites in this section are from GSC locality 94599, but elsewhere in the area the first trilobites occur slightly lower at about 30 feet (9.1 m) below the boundary between the Sekwi Formation and map unit 13 (ref. 8).

Fig. 3 Scanning electron micrographs of gold-palladium coated microfossils from GSC locality 94597. Each specimen is given a type number for storage in the National Type Fossil Collection, Geological Survey of Canada, Ottawa. *a-c*, *Protohertzina* cf. *anabarica* (GSC 45338). *a*, Posterior view; *b*, transverse section of distal end; *c*, transverse section of proximal end. *d-e*, Silicified concavo-convex microfossils of uncertain systematic position. *d*, Side view of GSC 45339; *e*, top view of GSC 45340. *f-h*, Microfossils, composed of agglutinated clay plates, possibly of foraminifer affinities; *f*, top view of GSC 45341; *g*, detail of exterior surface of GSC 45342 showing clay plates; *h*, side view of GSC 45343. Scale bars, 20 μ m (*g*); 100 μ m (*b*, *c*); 200 μ m (*a*, *d*, *e*, *f*, *h*).



end of an evolutionary series leading to the true or eucondonts⁴. (2) Approximately 12 fossils composed of agglutinated platy grains, apparently of a clay mineral (Fig. 3g). The fossils are roughly orthoconical and in basal outline are irregularly oval (Fig. 3f, h). Several phyla have representatives with agglutinated tests, but the small size of these fossils prompts comparison with protists and in particular the Foraminifera. The Lower Cambrian *Platysolenites* and *Spirosolenites*, regarded as the earliest agglutinated foraminifers, have a tubular form that is elongate in *Platysolenites* and coiled in *Spirosolenites*^{13,14}. The agglutinated foraminifer *Iridia*, which belongs to the primitive suborder Allogromiina¹⁵ has an irregularly hemispherical test, although with pseudochitinous base. As such a base is most unlikely to be fossilized, a comparison between the extant *Iridia* and these fossil forms may be valid (no intimate systematic relationship is at present claimed), and individuals of the latter may have lived adpressed to the substrate with pseudopodia extending from the margins. (3) About 200 silicified fossils with a pronounced concavo-convex shape (Fig. 3d, e). The degree of convexity varies, but hemispherical types predominate. In basal outline the specimens are usually slightly oval, but more elongate configurations are not uncommon. The lack of diagnostic shape or surface ornamentation precludes any useful speculation on their zoological affinities. This assemblage may well contain several different taxa. Use of very dilute acid (~1%) during rock digestion may produce more complete specimens.

Protohertzina anabarica has been recorded from the Siberian Platform (Anabar Shield) and Kazakhstan (Malij Karatau Ridge)¹⁰. In the north-west region of the Anabar Shield *P. anabarica* ranges through the latest Precambrian (?) Nemakit-Daldyn (= Manykay) Horizon¹⁰, and in Kazakhstan it occurs in equivalent strata¹⁶. The Nemakit-Daldyn Horizon, which has been correlated with the upper beds of the Yudoma Suite in the

classic Aldan River sections (refs 1, 2 and Rozanov in ref. 17), has a sparse shelly fauna which in addition to protoconodonts includes the worm tubes *Paleolina* and *Anabarites trisulcatus* (ref. 1 and Savitsky in ref. 17). A richer shelly fauna of Tommotian aspect has, however, been discovered close to the top of the Horizon near the type locality for *P. anabarica*¹⁰ on the Kotujkan River in the Anabar Shield¹⁸. Current Soviet usage^{1,2} generally places the Precambrian-Cambrian boundary at or very close to the first appearance of this richer fauna. Microfossils identified as *P. anabarica* have also been reported from the upper part (Huangshandong, Zhongyicun and Siowaitoushan members; the latter two formerly comprised the Meischucun Formation) of the Tongying Formation^{12,19} which is regarded as equivalent to the Tommotian of Siberia.

The presence of *P. cf. anabarica* in map unit 11 suggests that it is probably equivalent to part of the Nemakit-Daldyn Horizon or the lowermost Tommotian. Thus, according to current Soviet definition, the unit is close to the Precambrian-Cambrian boundary. If this correlation is correct the equivalents of the Tommotian Stage probably lie within map unit 12 and perhaps 13 (Fig. 2). The possibility remains that *P. cf. anabarica* has a higher stratigraphic range in the Mackenzie Mountains, but the great stratigraphic sequence (3,065 feet, 934 m) between *P. cf. anabarica* and the lowest Cambrian trilobites indicates otherwise.

Additional samples from the Mackenzie Mountains and elsewhere in the North American Cordillera are being processed, but the paucity of carbonate beds near the Precambrian-Cambrian boundary means that recovery of further microfossils may be limited. Recovery of other microfossils, such as acritarchs, through the use of stronger acids on clastic rocks has only begun and has so far met with limited success. Trace fossils in section 35 and nearby sections⁷ should be studied for evidence

of the pronounced diversification noted near the Precambrian–Cambrian boundary elsewhere in the world (see review in ref. 20). Preliminary field work at nearby section 33 (Fig. 1) provides grounds for optimism that this fossil diversification also exists in the Mackenzie Mountains (M. Fedonkin, personal communication). It is to be hoped that the diversification events in trace and shelly fossils can be tested against each other thereby improving intercontinental correlation of the Precambrian–Cambrian boundary.

We thank S. Bengtson, J. W. Cowie, G. S. Nowland and S. C. Matthews for helpful comments, J. Taylor, J. Hill, D. Newling and R. Carlton for technical assistance and P. Hammond and A. Mahoney typing. Initial research by S.C.M. was undertaken whilst a Research Fellow of St John's College, Cambridge. W.H.F. publishes with the permission of the Director of the Geological Survey of Canada.

Received 22 April; accepted 3 June 1980.

1. Matthews, S. C. & Missarzhevsky, V. V. *Q. Jl geol. Soc. Lond.* **131**, 289–304 (1975).
2. Rozanov, A. Yu *et al. Trudy Inst. geol. Nauk, Mosk.* **206**, 1–380 (1969).
3. Zhuravleva, I. T. *Am. J. Sci.* **269**, 417–445 (1970).
4. Bengtson, S. *Acta Univ. Upsal.* **415**, 1–71 (1977).
5. Rhoads, D. C. & Morse, J. W. *Lethaia* **4**, 413–428 (1971).
6. Stanley, S. M. *Paleobiology* **2**, 209–219 (1976).
7. Fritz, W. H. *Geol. Surv. Pap. Can.* **78–23**, 1–19 (1979).
8. Handfield, R. C. *Geol. Surv. Pap. Can.* **68–47**, 1–23 (1968).
9. Blusson, J. L. *Geol. Surv. Pap. Can.* **71–22**, 1–17 (1971).
10. Missarzhevsky, V. V. in *Palaeontological and Biostratigraphical Problems in the Lower Cambrian of Siberia and the Far East* (in Russian) 53–57 (1973).
11. Missarzhevsky, V. V. *Trans. Soviet–Mongolian palaeont. Expedn* **10–19** (1977).
12. Qian, Yi. *Acta paleont. sin.* **16**, 255–278 (1977).
13. Føyn, S. & Glaessner, M. F. *Norsk geol. Tidsskr.* **59**, 25–46 (1979).
14. Glaessner, M. F. *Bull. Bur. Miner. Resour. Geol. Geophys. Aust.* **192**, 61–65 (1978).
15. Loeblich, A. R. & Tappan, H. in *Treatise on Invertebrate Paleontology Part C* (Protista 2) (1964).
16. Mambetov, A. M. & Repina, L. N. in *Biostratigraphy and Palaeontology of the Lower Cambrian of Siberia* (in Russian), 98–138 (1979).
17. Cowie, J. W. & Glaessner, M. F. *Earth Sci. Rev.* **11**, 209–251 (1975).
18. Shishkin, B. B. *Geol. Geofiz. Novosibirsk* **4**, 111–114 (1974), (English translation in *Soviet geol. Geophys.* **15**, 97–99).
19. Qian, Yi, Chen, Menge & Chen, Yi-yuan *Acta paleont. sin.* **18**, 207–230 (1979).
20. Alpert, S. P. *Geol. Jl Spec. Issue* **9**, 1–8 (1977).

South-seeking magnetotactic bacteria in the Southern Hemisphere

R. P. Blakemore*, R. B. Frankel†
& Ad. J. Kalmijn‡

* Department of Microbiology, University of New Hampshire, Durham, New Hampshire 03824

† Francis Bitter National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

‡ Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Several species of aquatic bacteria which orient in the Earth's magnetic field and swim along magnetic field lines in a preferred direction (magnetotaxis) have been observed in marine and freshwater sediments of the Northern Hemisphere^{1,2}. Their orientation is due to one or more intracytoplasmic chains of single-domain magnetite particles³. These linearly arranged particles impart a net magnetic dipole moment to the bacterium, parallel to the axis of motility. Northern Hemisphere magnetotactic bacteria with unidirectional motility swim consistently in the direction of the magnetic field, that is, to the geomagnetic North^{1,2,4}. This implies that their magnetic dipole is systematically orientated with the North-seeking pole forward. The magnetic polarity can be reversed by single, magnetic pulses of high field strength (1–2 μ s, 300–600 G), and these bacteria then swim along magnetic field lines to the South⁵. Due to the inclination of the Earth's magnetic field, magnetotactic bacteria which swim to the North in the Northern Hemisphere are directed downward at an angle increasing with latitude. It has

been suggested that this downward-directed motion confers a biological advantage by guiding the bacteria, when dislodged, back to the sediments¹. On the basis of this hypothesis, magnetotactic bacteria of the Southern Hemisphere would be expected to swim to the South to reach the bottom. We report here several morphological types of magnetotactic bacteria present in sediments of the Southern Hemisphere. These bacteria indeed swim consistently to the South, hence downward along the Earth's inclined magnetic field lines, as hypothesized. As revealed by electron microscopy, they contain internal chains of electron-opaque particles similar to those observed in magnetotactic bacteria from the Northern Hemisphere. Like their Northern Hemisphere counterparts, their magnetic polarity can be permanently reversed and they cannot be demagnetized. We also report on Northern Hemisphere magnetotactic bacteria incubated in Southern Hemisphere magnetic conditions, confirming the biological relevance of downward directed motility.

Bacteria were collected from marine and freshwater sediments in the vicinities of Christchurch, New Zealand, and Hobart, Tasmania, Australia. These locales offered ecologically diverse habitats with geomagnetic characteristics (inclination 68–73°S at a total magnetic intensity of 0.59–0.63 G) approximately the mirror image of Northern Hemisphere environments where magnetotactic bacteria are found. Magnetotactic responses were recorded by optically monitoring bacterial motion in uniform magnetic fields of 0.2–2 G, provided by a pair of Helmholtz coils mounted to either side of a Zeiss OPMI 1 operation microscope (magnification $\times 83.2$; working distance, 10 cm). The Helmholtz-coil axis was aligned with the horizontal component of the Earth's magnetic field; the polarity of the imposed field was selected by a toggle switch. The coils and the dark-field illuminator were battery operated. In the coil system, both freshwater and marine bacteria migrated along magnetic field lines in the direction opposite to that indicated by a magnetic compass needle. On reversal of the field, the organisms executed U-turns and swam opposite to the initial direction. The rates of migration were comparable to those observed in Northern Hemisphere bacteria. Thus, the magnetotactic responses of Southern Hemisphere bacteria are similar to those of the Northern Hemisphere species, except for their magnetically opposite direction of migration (South-seeking magnetotactic bacteria have also been observed in Victoria, Australia⁶).

Transmission electron microscopy of bacteria magnetotactically separated from sediments was carried out at the University of Canterbury, Christchurch, New Zealand, using a Hitachi HS-7S electron microscope operating at 50 kV. Samples were placed on parlodion-coated, carbon-reinforced, 300-mesh copper grids, and were lightly stained with uranyl acetate. Magnetic cells of various morphological types were observed to contain one or two chains of intracytoplasmic electron-opaque particles which vary in size and shape with cell type (Fig. 1). Thus, Southern Hemisphere magnetotactic bacteria also possess magnetosomes⁷, the characteristic morphological feature of all magnetotactic bacteria examined thus far.

To establish the ferromagnetic nature of their orientation mechanism, bacteria were subjected to alternating 50-Hz magnetic fields over 1,000 G, produced by a small hand-held magnetic tape degausser (Tandy Corporation). Northern Hemisphere and Southern Hemisphere magnetotactic bacteria were examined in separate water drops on a microscope slide before and after exposure to the alternating magnetic field. On exposure, the degausser was slowly moved away from the water drops to obtain the field decay required for demagnetization. Whereas before exposure all the cells in these samples swam exclusively either to the North or to the South, after exposure each sample contained approximately equal numbers of North- and South-seeking cells. Magnetic reversals were also effected by approaching individual bacteria at the water–air interface with a sharp, magnetized needle. These observations indicate that bacteria from both hemispheres physically function as single-domain magnetic dipoles, in agreement with previous

remagnetization studies using strong magnetic pulses of short duration⁵.

In experiments carried out in Massachusetts, sediment samples containing North-seeking bacteria from New England were placed in an experimental chamber in which the vertical component of the local Earth's magnetic field was inverted, reproducing conditions in the Southern Hemisphere. Control samples were placed in a similar chamber with a field of normal inclination. The polarities of cells from the samples were determined periodically for several weeks, that is, over many generations. Whereas all samples contained almost exclusively North-seeking cells at the beginning of the experiment, with time the number of North-seeking cells in the experimental chamber decreased dramatically, and South-seeking bacteria gradually predominated. In the control samples no such changes occurred. Thus, downward-inclined fields select for North-seeking bacteria whereas upward-inclined fields select for South-seeking cells. This demonstrates that the vertical component of the Earth's magnetic field is the relevant parameter determining the predominant cell polarity.

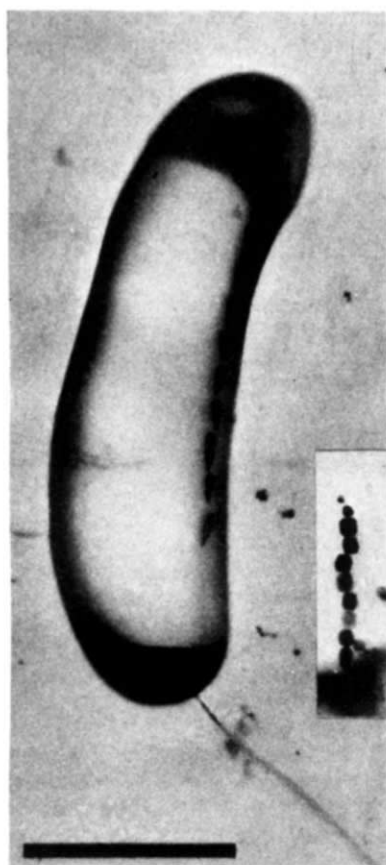


Fig. 1 Electron micrograph of a magnetotactic bacterium separated magnetically from freshwater sediments of New Zealand and stained with uranyl acetate. Inset shows one of two chains of particles found in a freshwater magnetotactic coccus of New Zealand. Both photographs are presented at the same magnification (scale bar = 1 μ m). Whereas the particles shown in the inset are similar to those found in Northern Hemisphere magnetotactic bacteria, those in the rod-shaped species have a novel shape.

In conclusion, the prevalence of South-seeking magnetotactic bacteria in Southern Hemisphere sediments and in samples held in similar magnetic conditions in the Northern Hemisphere verifies the hypothesis that downward-directed motion is advantageous for and upward-directed motion detrimental to the survival of these magnetotactic bacteria with unidirectional motility. Magnetotaxis is a reliable means of keeping these

microorganisms in or near the bottom sediments. As particles in the micrometre size range with densities close to 1.0 tend to remain suspended in water, gravity is virtually inconsequential in determining the vertical distribution of the bacteria.

The Earth's magnetic field provides a global orientation cue to which various organisms are known to respond. Only recently have detection mechanisms based on electromagnetic induction⁸ and ferromagnetic alignment^{5,9} been explained. In the bacteria, nature has shown the biological feasibility of synthesizing a highly organized ferromagnetic structure equivalent to a magnetic compass needle.

We thank Nancy Blakemore for her participation in the expedition to New Zealand and Tasmania, Australia, Drs T. Cole, L. Greenfield and W. D. Parkinson for hospitality and co-operation, M. Ingerfeld for assistance with electron microscopy, and H. Stram, D. Lobel, J. Pero, V. Kalmijn and T. Dourdeville for valuable contributions to this work. This work was supported by a research grant from the National Geographic Society, and the Office of Naval Research. The Francis Bitter National Magnet Laboratory is supported by the NSF and R.P.B. is supported, in part, by NSF grant PCM77-12175.

Received 14 February; accepted 9 May 1980.

1. Blakemore, R. P. *Science* **190**, 377 (1975).
2. Kalmijn, Ad. J. & Blakemore, R. P. *Proc. int. Un. physiol. Sci.* **13**, 364 (1977).
3. Frankel, R. B., Blakemore, R. P. & Wolfe, R. S. *Science* **203**, 355 (1979).
4. Moench, T. T. & Konetzka, W. A. *Archs Microbiol.* **119**, 203 (1978).
5. Kalmijn, Ad. J. & Blakemore, R. P. in *Animal Migration, Navigation and Homing* (eds Schmidt-Koenig, K. & Keeton, W. T.) 344 (Springer, New York, 1978).
6. Kirschvink, J. L. *J. exp. Biol.* (in the press).
7. Balkwill, D. L., Maratea, D. & Blakemore, R. P. *J. Bact.* **141**, 1399 (1980).
8. Kalmijn, Ad. J. *Handbook of Sensory Physiology* Vol. III (ed. Fessard, A.) 147-200 (Springer, Berlin, 1974).
9. Frankel, R. B. & Blakemore, R. P. *J. Magn. magn. Mater.* **15-18**, 1562 (1980).

Laetoli Pliocene hominid footprints and bipedalism

M. H. Day

Department of Anatomy, St Thomas's Hospital Medical School, London SE1 7EH, UK

E. H. Wickens

Department of Photogrammetry and Surveying, University College London, Gower Street, London WC1 EGBT, UK

The ability to stand upright and walk on two legs is widely recognized as a crucial hominid adaptation that had profound effects on the course of human evolution. The relief of the upper limb from the task of supporting the trunk in locomotion would have permitted the efficient carriage of food, infants and later tools and weapons. Any or all of these activities could, it is believed, have improved the differential survival of early hominids. Precisely when upright stance and bipedal gait were first acquired within the hominid evolutionary line is not known, and so the discovery of apparently bipedal footprints of Pliocene age at site G, Laetoli, Tanzania is important in itself^{1,2}. A study of the footprints, reported here, has shown that when these hominids walked, they transmitted their body weight and the forces of propulsion to the ground in a manner very similar to that of modern man.

Most of the evidence for early hominid upright posture and bipedal gait has been deduced from the anatomy of fossil hominid lower limb and pelvic bones from several African sites, including Sterkfontein and Swartkrans in the Republic of South Africa, Olduvai Gorge in Tanzania, Koobi Fora (formerly East Rudolf) in Kenya and the Afar site in Ethiopia³⁻¹¹. Opinions have varied as to the degree of adaptation to bipedalism that the

study of these fossil remains has disclosed¹²⁻¹⁷. The dates recorded for the sites where fossil hominid lower limb and pelvic bones have been found span from about 1.5 to 3.0 Myr BP, so that the discovery of what are believed to be bipedal footprints in 1978 and 1979 in a tuff dated at 3.6–3.75 Myr BP has added nearly 0.75 Myr to the known antiquity of bipedalism.

The fossil footprints of three individuals were made in a fine grained volcanic ash that subsequently became damp, perhaps after rain, and consolidated into a tuff retaining the imprints, often in considerable detail.

Because footprints made during walking reflect the way in which weight and propulsive forces are transmitted to the ground, we have compared the fossil footprints with those of modern man made experimentally on ground of similar consistency. Modern human footprints were made by walking over wet sand (grade 25/52, that is 0.30–0.6-mm grain size) in the laboratory so that clear impressions were made and retained for stereophotography. The instrument used was an Officene-Galileo stereometric camera on a 560-mm base, and the stereo images produced were contoured at 2-mm intervals on a Thompson-Watts Mk 2 plotting instrument. This technique—photogrammetry—is widely used by surveyors and map-makers as a precision method of determining surface form. The

experimental footprint contour diagrams (Fig. 1) show the various depths that resulted from the differing pressures exerted on the ground in walking; the darker the colour, the deeper the impression. Similar diagrams were prepared from stereophotographs of replicas of three fossil footprints from two individuals (Figs 2, 3).

The remarkable similarities between the modern human footprint contour patterns and those of the fossil hominids can be seen at a glance. In more detail, heel and ball impressions are clear and transfer of weight and force progresses from heel to ball on the outer side of the foot outlining a medial arch. The heel and ball impressions are deep and surrounded by an irregular figure-of-eight pattern in each case. The ball impressions have oblique principal axes that run anteromedially towards the base of the great toe, where the impressions are especially deep, and lead on to the great toe impression itself as it digs into the soft ground during walking.

The similarity of the contour patterns of the three fossil footprints to the modern human examples is striking confirmation of the bipedal nature of the gait of the Laetoli Site G hominids already claimed from the footfall sequence^{1,2}. The pattern of weight and force transference through the foot, well known in modern man¹⁸, also seems to be very similar in the



Fig. 1 Modern human female (a) and male (b) soft ground walking footprints. Contoured at 2-mm intervals.



Fig. 2 A right fossil hominid footprint from site G, Laetoli. The footprint shows many points of similarity with those of modern man. Contoured at 2-mm intervals.

fossil footprints and indicates that even at this early stage of hominid evolution bipedalism had reached an advanced and specialized stage. Ironically these fossil hominid footprints from Laetoli provide locomotor evidence without a single adult fossil hominid limb bone being known from the site. Nonetheless Pliocene hominid footprints, and their comparison with experimental modern human equivalents, may well prove to be more convincing evidence of early hominid bipedalism than osteological or morphometric analysis of fossil bones.

We thank Dr Mary Leakey for the opportunity to examine the Laetoli footprints; Mr A. A. Mturi, Director of Antiquities, Tanzania, Mr A. J. N. Mgina, Ngorongoro Conservation Authority, for help and cooperation; the Research Endowments Committee at St Thomas's Hospital for support, and colleagues in the Department of Photogrammetry and Surveying, University College London.

Received 10 December 1979; accepted 3 June 1980.

1. Leakey, M. D. *Antiquity* **52**, 133 (1978).
2. Leakey, M. D. & Hay, R. L. *Nature* **278**, 317 (1979).
3. Broom, R., Robinson, J. T. & Schepers, G. W. H. *Trans. Mus. Mem.* **4** (1950).
4. Napier, J. R. *Arch. Biol. Liège* **75**, Suppl. 673 (1964).
5. Davis, P. R. *Nature* **201**, 967 (1964).
6. Day, M. H. & Napier, J. R. *Nature* **201**, 967 (1964).
7. Day, M. H. & Napier, J. R. *Nature* **211**, 929 (1966).
8. Day, M. H. *Nature* **215**, 323 (1967).
9. Day, M. H. *Nature* **221**, 230 (1969).
10. Day, M. H. in *Earliest man and the Environment of the Lake Rudolf Basin* (eds Coppens, Y., Howell, F. Clark, Isaac, Glynn L.I. & Leakey, R. E. F.) (Chicago University Press, 1976).
11. Johanson, D. C. & Taib, M. *Nature* **260**, 293 (1976).
12. Dart, R. A. *Am. J. Phys. Anthropol.* **7**, 301 (1949).
13. Clark, W. E. Le Gros *The Fossil Evidence for Human Evolution* (The University of Chicago Press, 1955).
14. Robinson, J. T. *Early Hominid Posture and Locomotion* (Chicago University Press, 1972).
15. Day, M. H. *Symp. zool. Soc.* **33**, 29 (1972).
16. Zuckerman, S. *et al. Symp. zool. Soc.* **33**, 71 (1972).
17. Lovejoy, C. O. & Heiple, K. G. *Nature* **235**, 175 (1972).
18. Barnett, C. H. *Lancet* **i**, 617 (1956).

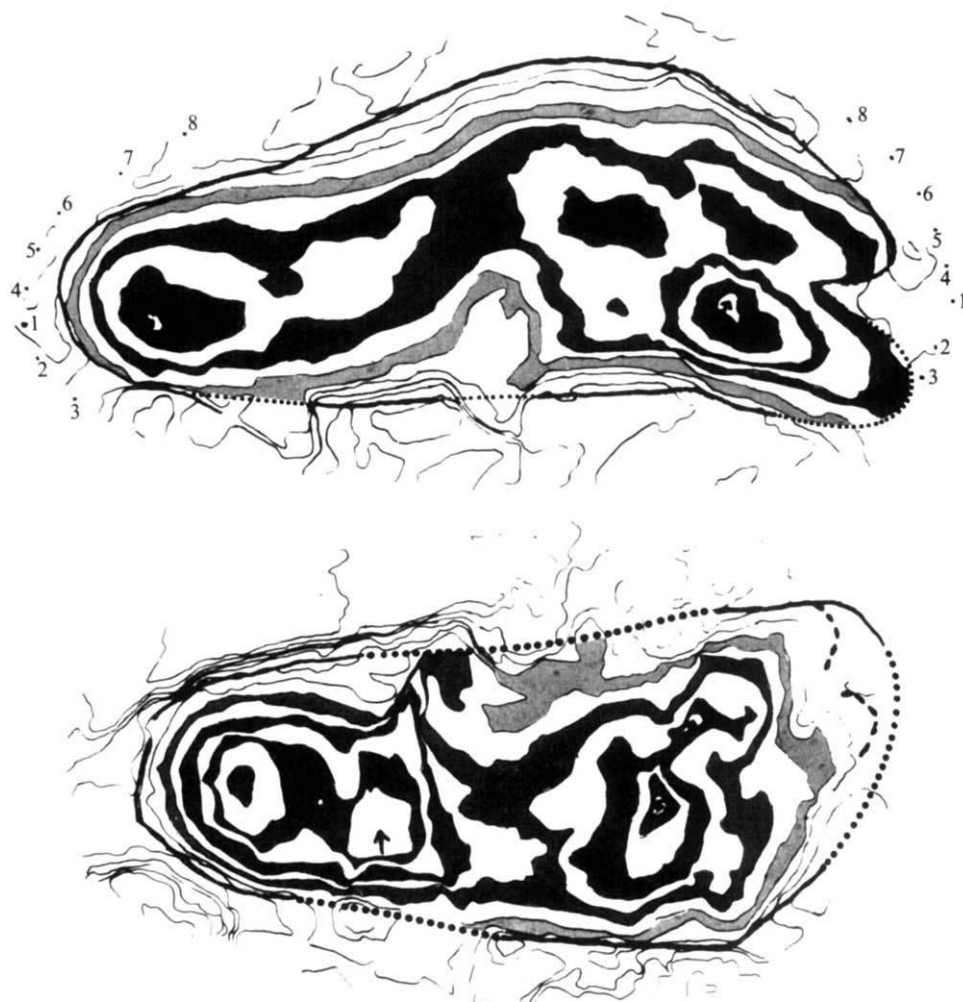


Fig. 3 Left (a) and right (b) fossil hominid footprints from a larger individual that also show striking similarities to those of modern man. Contoured at 2-mm intervals.

Environmental preference induced experimentally in ants (Hymenoptera: Formicidae)

Pierre Jaisson

Equipe de Recherche Associée au CNRS no. 885, Laboratoire d'Éthologie et Sociobiologie, Université Paris-Nord, Avenue J.B. Clément, 93430 Villetaneuse, France

In ants, early experience seems to influence the development of aggression^{1,2} and the recognition of cocoons^{3,6}, but there has been no investigation of such influences on the ants' relationship with their habitat. There are, however, many descriptions of the use of plants as nests (*Azteca*⁷, *Colobopsis*⁸, *Oecophylla*⁹ and *Pseudomyrmex*¹⁰, for example), while habitat selection based on early experience is known among some vertebrates, principally amphibians¹¹⁻¹³, birds¹⁴ and mammals¹⁵. In this context, my hypothesis is that the preference of an ant for a plant can be induced on the basis of previous entrainment to a plant stimulus. Using ants of the sub-family Formicidae, *Camponotus vagus* Scop. and *Formica polyctena* Först., I have demonstrated that early experience can induce an environmental preference and may explain specific associations that can be found in the wild between ants and plants.

C. vagus and *F. polyctena* seem to have close affinities with plants. The former lives inside wood, while the latter builds nests of twigs or pine needles. The two species are common in France, and colonies were collected either in Provence (near Vaison-la-Romaine, Vaucluse) in the case of *C. vagus*, or the Rambouillet forest (Yvelines). Laboratory colonies were established using very young ants which all emerged on the same day and which were placed in a glass tube (22×2 cm) with a source of water at the closed end, providing the necessary humidity; an envelope of black card obscured the light.

Colonies comprised 25–45 individuals, depending on the number of ants eclosing weekly. Three 5-cm sprigs of thyme (*Thymus vulgaris* L.) were placed near the water reservoir in the experimental colonies. Control colonies had no thyme. Sprigs were replenished each week during the training period which lasted 30 days for *C. vagus* and 21 days for *F. polyctena*. Food consisted of honey and water. All colonies were then transferred to tubes without thyme for 24 h; then, anaesthetized with carbon dioxide, each colony was put into a plastic box equidistant from two lateral horizontal plastic tubes fixed into the sides, one of which was randomly chosen to contain thyme (to avoid any effect of laterality).

Each tube was a potential nest. The apparatus was illuminated isotropically, with white light, so that the interior of each tube (covered by an envelope of black card) was relatively dark. On recovery from anaesthesia, the ants, avoiding the light, explored the box and finally chose the shade in one of the tubes. On the following day the ants in each tube were counted. This choice test was applied to 15 control and 9 experimental colonies of *C. vagus*, and 9 control and 13 experimental colonies of *F. polyctena*. The few ants remaining in the box were also counted but not included in the determination of colony preference. This preference was evaluated as the difference between the number of ants that chose the nest with thyme and the number that chose the nest without thyme.

Table 1 shows that all colonies of *C. vagus* trained to the thyme stimulus chose the tube containing thyme. Conversely, all but one of those familiarized to an environment without thyme preferred the tube without thyme. But even this colony, although found in the tube that had contained thyme at the beginning of the test, moved all the sprigs into the plastic box. Nevertheless, I decided that this colony had chosen the thyme

Table 1 Behaviour of young *C. vagus* and *F. polyctena*

		N	T+	T−	d
<i>C. vagus</i>	Experimental colonies	9	171	41	9
	Control colonies	15	72	330	1
<i>F. polyctena</i>	Experimental colonies	13	348	33	11
	Control colonies	9	89	181	2

Experimental colonies comprised 25–30 workers, and control colonies 30 workers for *C. vagus*; experimental and control colonies comprised 25–45 workers for *F. polyctena*. N, number of colonies; T+, ants that preferred the tube with thyme for all colonies; T−, ants that preferred the tube without thyme; d, colonies in which T+ > T−.

side. In fact, the two tubes were empty when the count was made.

I used non-parametric statistical analysis: at the colony level, the χ^2 total correlation is highly significant ($\chi^2=16.5$; $P<0.01$). This analysis was applied to all the colonies to test the nul hypothesis of a randomized distribution of the colonies between two possibilities (preference for thyme/preference for empty tube). The same test applied only to each group as a whole (experimental or control) confirms my first observations: in both cases the probability that the results represent a randomized distribution of the colonies during the choice test is less than 0.01. Thus the preference of the experimental group for the nest containing thyme is significant ($\chi^2=8.8$), as is that of the control group for the empty nest ($\chi^2=11.3$).

Wilcoxon's test gave the same results, but more precisely than χ^2 . It enabled me to classify the colonies on the basis of the difference between the number of ants in the colony that preferred the tube with thyme and the number in the colony that preferred the empty tube. In other words, it enabled me to associate the qualitative and quantitative aspects of the preference (the tables show only the total group values).

In terms of the number of ants involved (namely their individual preference), χ^2 , either total or calculated for each of the two groups, is highly significant, regardless of the fact that the colonies were artificially set up. However, there is a need for caution because each ant's choice between the two artificial nests is not standard: each new entry into a tube may be qualitatively different from the preceding entry.

The same applies to *F. polyctena*, as Table 1 shows, although after choosing the tube containing thyme, the two control colonies did not expel the sprigs. The same statistical treatment shows again that the behaviour of the colonies as a whole was heterogeneous ($\chi^2=6.19$; $P<0.02$) and the experimental colonies had a significant preference for the tube containing thyme ($\chi^2=6.23$; $P<0.02$). But the same test applied to control colonies gives a non-significant value ($\chi^2=1.77$). Wilcoxon's test does not improve the precision (only significant for the experimental group). However, the comparison of individual ant preference for the tube without thyme in the controls becomes significant ($\chi^2=31.3$; $P<0.01$); nevertheless, it should be remembered that treatment of the individual results increases the information but modifies somewhat the significance of the choice test.

Thus, the behaviour of these two species differed according to previous experience: the experimental colonies, trained to thyme, chose the tube containing it, while the control colonies preferred the tube without thyme. However, these results were more clear-cut for *C. vagus*.

To test the possible importance of age, I examined environmental preference in older ants, collected from nests in the same localities as before. These experiments were carried out at the end of the season, when I could form only seven

colonies of *C. vagus* and five of *F. polyctena*, to which the same training period and bioassay was applied. Table 2 shows, first, that adult ants reared in the presence of thyme suffered heavy mortality: 48.6% for *C. vagus* and 56.6% for *F. polyctena* (compared with 9.8% and 5.8% for the young ants). It seems to be deleterious for naive adult ants to be exposed to this plant, for which they have a spontaneous aversion (as known already in insects¹⁶). The aversion was revealed by the distribution of the workers, especially at the beginning of the training period: most of them remained close to the sprigs, rather than within the thyme as the young workers had done.

In the choice test, statistical analysis revealed responses much less clear than for the young ants. Preference for the empty tube seemed to be clear-cut for *C. vagus* ($\chi^2 = 5.14$; $0.05 > P > 0.01$), but not significant for *F. polyctena* ($\chi^2 = 5.26$; $0.05 > P > 0.01$).

Adult ants trained to the thyme always seemed to prefer a tube without thyme, even if this preference (in particular for *F. polyctena*) appeared less pronounced than for control colonies. Although the high mortality of the adults exposed to thyme affected the number of individuals, the sensitivity of the young workers seemed to be greater and their age more favourable for the establishment of an environmental preference.

The behaviour of workers in the presence of thyme varied according to age. Mature ants avoided contact with the plant more often than not, while the young ants entered the sprigs. This difference may enable the latter to develop an active preference for thyme on the basis of its presence during the neonatal period.

Table 2 Behaviour of mature *C. vagus* and *F. polyctena*

	N	T+	T-	d
<i>C. vagus</i>	7	11	115	0
<i>F. polyctena</i>	5	28	48	2

In each case there were 35 worker ants.

There have been reports of avoidance conditioning in mature ants for the odour of peppermint (used as a conditional stimulus)^{17,18}. But the preference reported here could resemble more the olfactory conditioning discovered by Thorpe and Jones¹⁹, as in two species of *Drosophila*^{20,21}, which if reared in a culture medium scented with peppermint oil, later showed a positive response to that odour. However, that phenomenon, which is difficult to interpret²², develops only when peppermint is presented simultaneously with food during training²³. Conversely, in my experiments food was presented very briefly at the nest entrance, away from the thyme.

I suggest that the induction of the preference for thyme in young ants is due to either passive familiarization or an association between favourable environmental factors (for example darkness and/or humidity in my experiments) and the odour of the plant. The repulsive nature of plant odour for young ants has not been established, and their sensory receptors might increase progressively in sensitivity as the ants develop, as shown for the phytophagous larva of *Manduca sexta* reared on a plant that it normally finds unacceptable²⁴. A mechanism of this type, but independent of feeding, may contribute to the development of relationships of the sort involving the ants *Azteca* or *Pseudomyrmex*, which live exclusively on the plants *Cecropia* or *Acacia*, respectively.

Received 20 August 1979; accepted 25 April 1980.

1. Fielde, A. M. *Biol. Bull.* **10**, 1 (1902).
2. Fielde, A. M. *Indep. Aug.* **17**, 375 (1905).
3. Jaisson, P. *C.r. hebd. Séanc. Acad. Sci. Paris* **275**, 2721 (1972).
4. Jaisson, P. *Behaviour* **52**, 1 (1975).
5. Le Moli, F. & Passetti, M. *Att. Soc. Ital. Sci. Nat., Milano* **118**, 49 (1977).
6. Jaisson, P. & Fresneau, D. *Anim. Behav.* **26**, 1064 (1978).
7. Janzen, D. H. *Ecology* **50**, 147 (1969).
8. Wheeler, W. M. *Ants: Their Structure, Development and Behavior* (Columbia University Press, New York, 1910).

9. Way, M. J. *Bull. ent. Res.* **45**, 93 (1954).
10. Janzen, D. H. *Kans. Univ. Sci. Bull.* **47**, 315 (1967).
11. Wiens, J. A. *Copeia* **1970**, 543 (1970).
12. Wiens, J. A. *Anim. Behav.* **20**, 218 (1972).
13. Mushinsky, H. R. *Copeia* **1976**, 755 (1976).
14. Hildén, O. *Ann. zool. Fenn.* **2**, 53 (1965).
15. Wecker, S. C. *Ecol. Monogr.* **33**, 307 (1963).
16. Soliman, M. H. & Hay, D. A. *Experientia* **34**, 329 (1978).
17. Hoagland, H. J. *gen. Psychol.* **5**, 21 (1931).
18. Vowles, D. M. J. *comp. Physiol. Psychol.* **58**, 105 (1964).
19. Thorpe, W. H. & Jones, F. G. W. *Proc. R. Soc. B124*, 56 (1937).
20. Thorpe, W. H. *Proc. R. Soc. B127*, 424 (1939).
21. Cushing, J. E. *Proc. natn. Acad. Sci. U.S.A.* **50**, 163 (1941).
22. Alloway, T. M. A. *Rev. Ent.* **17**, 43 (1972).
23. Hershberger, W. A. & Smith, M. P. *Anim. Behav.* **15**, 259 (1967).
24. Schoonhoven, L. M. *Proc. Koninkl. Ned. Akad. Wetens., Amsterdam* **72**, 491 (1969).

Orientation bias of cat retinal ganglion cells

W. R. Levick & L. N. Thibos

Department of Physiology, John Curtin School of Medical Research, Canberra, Australia 2601

Kuffler¹ described the receptive fields of cat retinal ganglion cells as having a concentric arrangement. This has usually been taken to mean that they are approximately circular in form (see, for example, ref. 2). Hammond³ tested the circularity of the centre component of receptive fields by plotting a contour of isosensitivity to a small flashed spot. He concluded that centres were often elliptical (average ratio of major to minor axis 1.23) and that more than 50% of the recorded cells had the major axis oriented within $\pm 20^\circ$ of the horizontal. Such data are important for discussions⁴ of the neurophysiological basis of the 'oblique effect' (reduced visibility for periodic grating patterns when oriented away from the vertical or horizontal) observed in psychophysical experiments on humans because subcortical units are often assumed to be orientationally unbiased. Orientation selectivity is a prominent attribute of visual cortical neurones⁵ so analysis has usually emphasized the distribution of orientation selectivity at that level⁶⁻⁸. The results presented here redirect attention to the retinal level since they reveal a previously unsuspected systematic relation between orientation bias of ganglion cells and their location relative to the area centralis.

Experiments were conducted on adult cats anaesthetized for surgical procedures with 1-2% Fluothane and maintained during data collection on a 70:28:5:1:5 mixture of nitrous oxide, oxygen and carbon dioxide, usually supplemented by intermittently administered urethane (about 0.6 g per day). Eye movements were controlled by continuous infusion of gallamine (5 mg per kg per h) and tubocurarine (0.7 mg per kg per h). The paralysis necessitated artificial ventilation. Extracellular recordings were made from ganglion cells with a tungsten-in-glass electrode⁹ passed through a cannula penetrating the globe so as to leave the natural optical system undisturbed. Because optical quality of the retinal image may influence results special precautions were taken as follows. Corneal refraction was modified with a plastic contact lens of spherical form and zero power. The resulting ocular refraction was measured with a Hartinger refractionometer. The point of passage of the axis of the instrument through the dilated natural pupil (1% atropine drops) was adjusted so as to minimize astigmatism. An artificial pupil (3 mm diameter) was centred on this axis and close (about 1 mm) to the contact lens. The combination of spherical and cylindrical lenses indicated by the refractionometer was placed next to the pupil and further measurements were made to confirm the emmetropic state.

The stimulus was a sine wave grating (contrast 0.5, mean luminance 82 cd m⁻²) drifting slowly (2 bars per s passing a given point) across the receptive field. The pattern was generated as a raster on the face of a cathode ray tube (Tektronix 608) which

was centred on each receptive field in turn by means of a gimbal mounting. The orientation of the bars of the grating was controlled by a Dove prism also carried on the gimbal. The visual subtense of the pattern through the prism was about 20° . The screen was focussed at optical infinity with a +4 dioptre lens placed close to the prism between screen and prism. A computer (Hewlett Packard 2100) controlled both the production of the raster and the measurement of trains of impulses constituting the responses of a ganglion cell. The instant when a particular phase of each cycle of the grating passed the centre of the screen provided the synchronization for accumulating usually 100 cycles of response as a 'peri-stimulus time histogram' (PSTH).

As the spatial frequency of a grating was increased over the range 0.03–8 cycles per degree, modulation of a ganglion cell's discharge first increased to a maximum, then decreased. By working near the high-frequency limit for a modulated response from each ganglion cell it was easy to find instances where response modulation was strongly influenced by grating orientation. The orientations yielding maximum and minimum modulation were generally 90° apart. Two obvious examples are shown in Fig. 1. A convenient measure of response is the amplitude of the first harmonic component obtained by Fourier analysis of the PSTH. In the examples of Fig. 1 the ratio of maximum to minimum response was 3.2 for unit 1 and 10.7 for unit 2. In a sample of 250 units from five cats the distribution of the ratio ranged from 1.0 to 18.7 with the median at 1.86. By extrapolating the relations between response and spatial frequency to determine the spatial frequency at zero response ('cutoff frequency') for best and worst orientations, the behaviour may also be expressed as a ratio of cutoff spatial frequencies. Ratios of up to 1.6 have been encountered. Clearly the orientation dependence is a definite effect. However, the dependence was insignificant in 30% of the cells tested and this may explain why it was not reported previously².

In several animals where data were collected from many ganglion cells in a local patch of retina, it became evident that

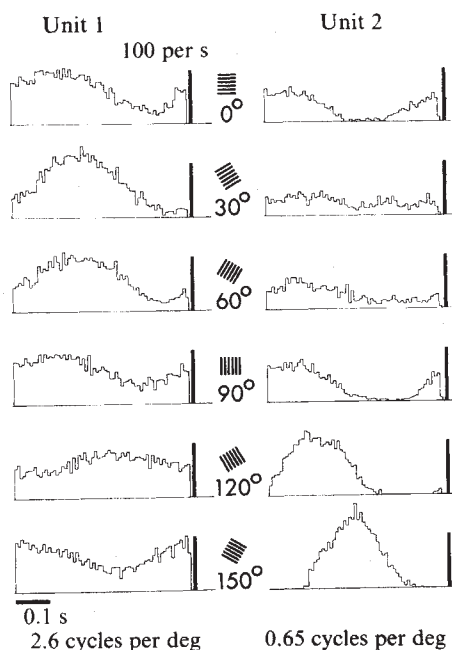


Fig. 1 Peristimulus time histograms (8 ms bin width) of responses to a sinusoidal grating (contrast 0.5) at various orientations (indicated schematically down centre column) drifting at a constant rate (2 c.p.s.) across the receptive field of two ganglion cells: left, on-centre brisk-sustained type; right, on-centre brisk-transient. Same vertical scale (solid bar at right of each, 100 impulses per s) for all PSTHs. Spatial frequencies indicated at bottom are just below the highest resolvable by each unit. The preferred orientations were judged to be 30° , 150° respectively. The positions of the cells relative to the area centralis are indicated (1, 2) in Fig. 2.

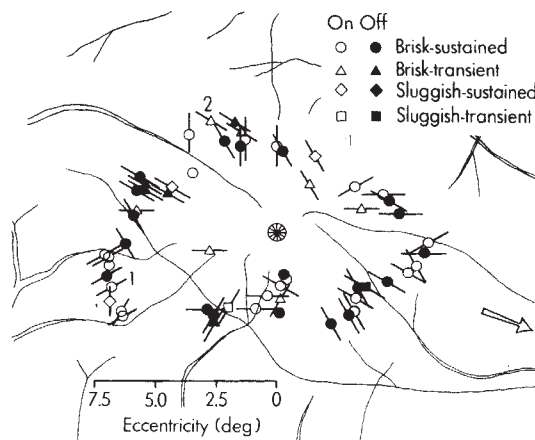


Fig. 2 Preferred grating orientation in relation to retinal location. Positions of receptive fields are indicated by symbols (keyed to unit class, top right) on a tracing of a fundus photograph showing pattern of blood vessels and centre of area centralis (spoked wheel symbol). A line through a symbol indicates the preferred orientation of that unit (no line—no orientation bias). The two units of Fig. 1 are distinguished: 1, 2. The arrow points along the line from area centralis to optic disk.

there was a clustering of best orientations. Patches from various locations had different best orientations. The central tendency of best orientations seemed to be consistently parallel to the line joining the patch to the area centralis. The uncertainties of combining diverse data from various animals influenced us to examine a substantial number of ganglion cells in a single preparation. A sequence of recordings was made along an approximately circular trajectory completely encircling the area centralis. Each unit was assessed as in Fig. 1 and the best orientation recorded as that yielding the maximum modulated response. The position of the centre of each receptive field was transferred to a fundus photograph with the aid of three independent assessments: (1) direct ophthalmoscopic visualization of the electrode tip relative to the pattern of retinal blood vessels; (2) hand mapping of receptive field centres on a frontal tangent screen; (3) noting azimuth and elevation readings on the gimbal carrying the centred cathode ray tube. The photograph was also marked with the centre of the area centralis estimated from the peak of ganglion cell density in the cresyl violet stained whole-mount of the retina and correlation of the blood vessel pattern in whole-mount and photograph.

The result is displayed in Fig. 2 which is a tracing of the fundus photograph. Out of 57 units examined, 52 showed definite orientation bias. Their positions and best orientations are indicated. It is evident in a statistical sense that orientation bias was systematically related to retinal locus. A similar result was obtained in another experiment of the same design. The effect was observed in all classes of concentrically organized ganglion cells¹⁰.

A particular point of interest in these results is that they indicate retinal regions where the information sent to the brain is biased in favour of obliquely oriented gratings. As far as we are aware, eccentric regions along oblique meridians have not been systematically explored in human psychophysical experiments.

Received 4 February; accepted 2 May 1980.

1. Kuffler, S. W. *Cold Spring Harb. Symp. quant. Biol.* **17**, 281–292 (1952); *J. Neurophysiol.* **16**, 37–68 (1953).
2. Enroth-Cugell, C. & Robson, J. G. *J. Physiol., Lond.* **187**, 517–552 (1966).
3. Hammond, P. W. *J. Physiol., Lond.* **234**, 64P–66P (1973); **242**, 99–118 (1974).
4. Appelle, S. *Psychol. Bull.* **78**, 266–278 (1972).
5. Hubel, D. W. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106–154 (1962).
6. Mansfield, R. J. W. *Science* **186**, 1133–1135 (1974); *Brain Res.* **149**, 229–234 (1978).
7. Rose, D. & Blakemore, C. *Expl Brain Res.* **20**, 1–17 (1974).
8. Finlay, B. L., Schiller, P. H. & Volman, S. F. *Brain Res.* **105**, 350–352 (1976).
9. Levick, W. R. *Med. Electron Biol. Engng* **10**, 510–515 (1972).
10. Cleland, B. G. & Levick, W. R. *J. Physiol., Lond.* **240**, 421–456 (1974).

Voltage clamp discloses slow inward current in hippocampal burst-firing neurones

Daniel Johnston & John J. Hablitz

Program in Neuroscience and Section of Neurophysiology, Department of Neurology, Baylor College of Medicine, Houston, Texas 77030

Wilkie A. Wilson

Epilepsy Center, Veterans Administration Hospital, and Department of Pharmacology, Duke University Medical Center, Durham, North Carolina 27705

Hippocampal pyramidal cells, among the most-studied cortical neurones, display several interesting properties. In particular, intracellular recordings in both *in vivo*¹ and *in vitro*² preparations have shown that the pyramidal neurones of the CA3 region of the hippocampus have, in addition to the more conventional single spike, the capacity to generate spontaneous bursts of action potentials (APs). The burst-firing patterns have been well documented, but their underlying electrophysiological mechanisms are not completely understood. The bursts seem to be an intrinsic response of individual neurones and can be triggered by weak orthodromic stimulation or by direct depolarization of the somatic membrane. In molluscan and spinal systems, voltage-clamp techniques have revealed slow, voltage-sensitive, inward currents underlying this type of firing behaviour³⁻⁷. The use of these voltage-clamp techniques to study this phenomenon in hippocampal neurones has previously been precluded by the small size of the neurones and by the difficulty of recording from them in intact preparations. However, the development of methodology for voltage-clamping with a single microelectrode (SEC)⁸ and for the maintenance of hippocampal slices *in vitro*⁹ has made it possible for us to examine the membrane currents in these CA3 neurones. We report here a slow inward current in CA3 neurones and suggest that it is carried by calcium ions.

Experiments were carried out on hippocampal slices from guinea pigs weighing 300–700 g. Slices 400–600 μm thick were cut transverse to the longitudinal axis of the hippocampus and maintained *in vitro*, using standard techniques^{9,10}. Intracellular recordings were obtained from CA3 neurones, using 3 M KCl- or 2 M CsCl-filled micropipettes (30–50 M Ω). Electrodes were selected carefully for current-passing capability and for rapid decay of tip potentials following cessation of current.

The circuit used for voltage clamping is based on that of Wilson and Goldner^{8,11}. The technique uses a single microelectrode and rapid switching (3,000 Hz) between voltage-recording and current-passing modes. The membrane potential is sampled after the potential across the microelectrode has decayed and just before return to the current-passing mode. The sampled potential is fed into a standard voltage-clamp circuit and determines the amplitude of the next current pulse. For excursions of approximately ± 30 mV from rest, a steady potential could be reached within 2 ms.

A standard voltage-clamp protocol was used, whereby step depolarizing commands were given to the cell while the current response was monitored. Figure 1a shows the response of a CA3 neurone, bathed in saline containing $1 \mu\text{g ml}^{-1}$ tetrodotoxin (TTX), to 300-ms commands. At small depolarizations, only the leakage current is seen; at larger depolarizations, a slow, inward-going current develops and peaks in about 50–100 ms. Either the inward-going current decays during the command, or an outward current develops and reverses the inward-going trend. The total current-voltage (*I-V*) relationship of this cell in

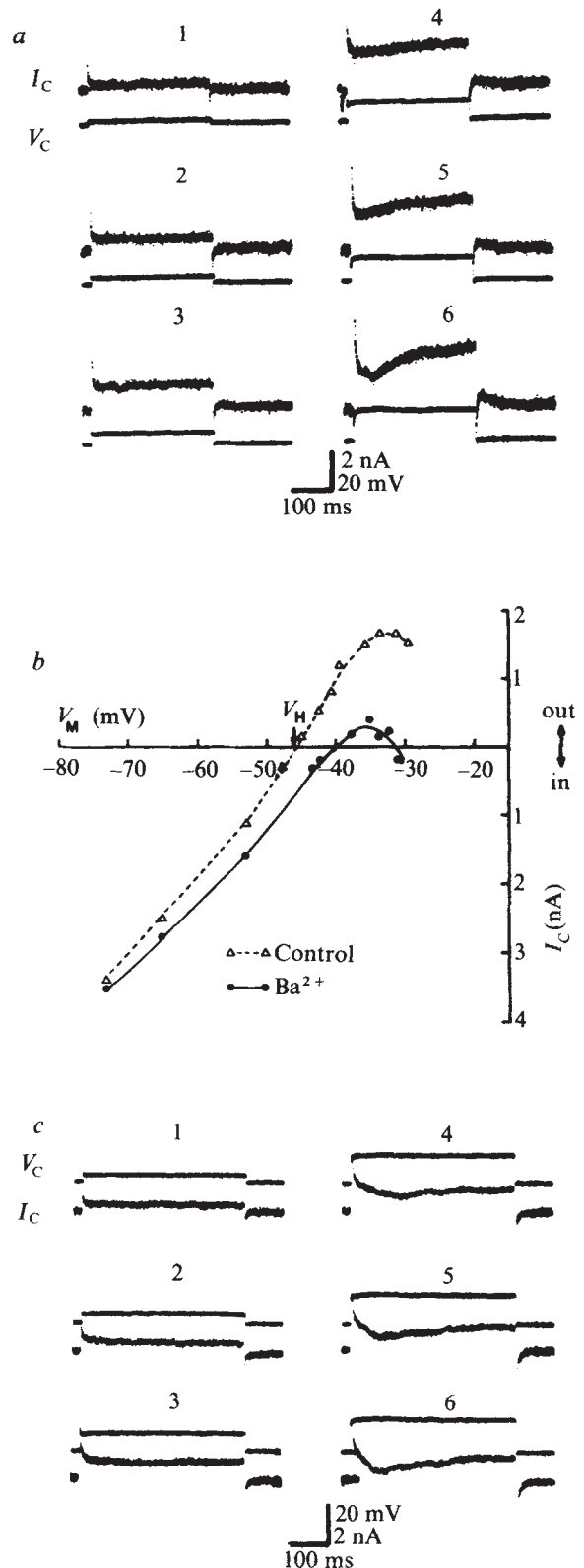


Fig. 1 Voltage-clamp data from a CA3 neurone in the presence of $1 \mu\text{g ml}^{-1}$ TTX. *a*, A series of depolarizing commands is given to the cell. Frames 1–3 show primarily leakage current. In frames 4–6, an inward-going current develops that peaks in 50–100 ms. *I_c*, clamp current; *V_c*, command potential. (See 1*b* for value of holding current.) *b*, *I-V* curve from same cell before and after applying a drop of 15 mM Ba^{2+} . Current is measured as the peak of the inward-going phase. *V_M*, membrane potential under voltage clamp; *V_H*, holding potential. After application of Ba^{2+} , the *I-V* curve becomes net inward at about -30 mV. *c*, Depolarizing clamp commands after applying Ba^{2+} .

normal saline plus TTX is shown in Fig. 1b; the current is measured with respect to zero at the peak of the inward-going phase. A similar inward-going current has been observed in well over 50 CA3 neurones. In Fig. 1c, current from depolarizing commands is shown from the same cell after applying a drop of 15 mM Ba^{2+} from a coarse extracellular micropipette. It is clear that after the application of Ba^{2+} the total membrane current is less outward at all potentials, membrane current becomes net inward at about -40 mV (Fig. 1b), and the inward-going current is more prominent during the depolarizing commands (Fig. 1c). Since it is likely that the depolarizing commands activate both an outward K^+ current and a TTX-insensitive inward current simultaneously, and as Ba^{2+} has been shown, in several preparations, to block K^+ currents^{12,13} as well as to substitute for Ca^{2+} as the charge carrier through Ca^{2+} channels^{14,15}, further manipulations were necessary to separate the inward and outward currents.

Internally applied caesium (Cs^+) has been shown to block outward K^+ currents in a variety of systems¹⁶, and we used Cs^+ loading to reduce the K^+ contamination of the inward current in CA3 neurones. Neurones were impaled with micropipettes containing 2 M CsCl , and Cs^+ was iontophoresed for 1–10 min using 200–400-ms pulses at 2 Hz. Cs^+ injection typically increased the passive input resistance two- or threefold and seemed to block K^+ currents, as judged by: (1) the increased duration of the APs; (2) the decreased leakage conductance; and (3) the decreased outward current observed in voltage clamp. Moreover, the passive input resistance, as measured under current clamp, showed little decrease with depolarization up to about -40 mV, suggesting a substantial reduction in delayed rectification. Figure 2 shows an I - V relationship from a Cs^+ -injected neurone in which Na^+ currents had previously been blocked by TTX. In this cell, the total membrane current was inward over the entire voltage range up to -10 mV, the limit to which we could clamp the cell reliably; a clear region of negative slope conductance in the range of -45 to -20 mV is demonstrated.

To test for the ionic sensitivity of this slow inward current, Co^{2+} , a potent inhibitor of Ca^{2+} inward currents in a variety of preparations^{14,15}, was applied to CA3 neurones. Figure 3a shows data from another Cs^+ -injected CA3 neurone in a TTX-treated slice. Although in this cell the holding current is outward at the holding potential of -40 mV (Fig. 3b), the depolarizing commands activate a persistent inward-going current (see Fig. 3a, 3 and 4) that becomes net inward at about -25 mV. Hyperpolarizing commands from -40 mV show little leakage current (Fig. 3a, 1 and 2). A drop of 10 mM Co^{2+} was applied to this cell via a coarse extracellular micropipette. Figure 3c shows that following this application, the current response to step commands is quite different. No inward-going current is activated by

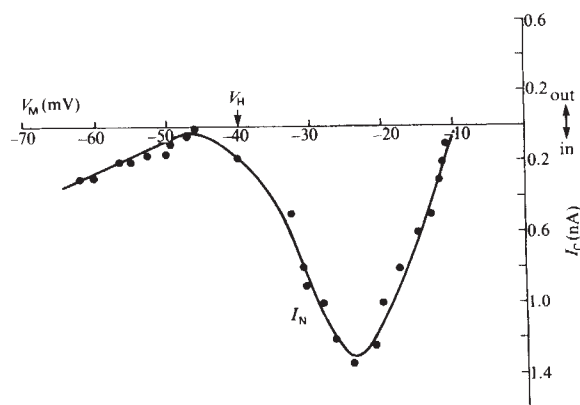


Fig. 2 I - V curve of a CA3 neurone, injected with Cs^+ . The cell was treated with $1 \mu\text{g ml}^{-1}$ TTX, and Cs^+ was injected via the intracellular microelectrode. Total current is inward at all potentials. A prominent region of negative slope conductance is seen after Cs^+ loading.

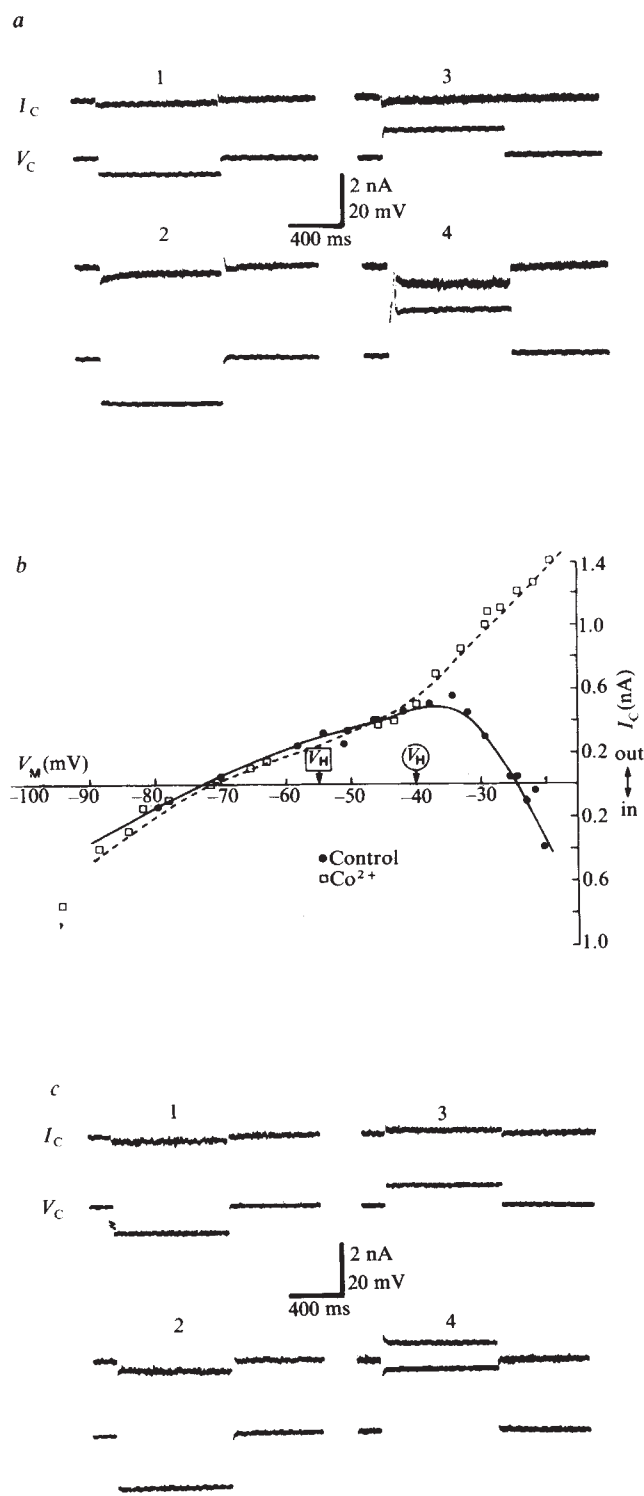


Fig. 3 Effects of Co^{2+} on a CA3 neurone. *a*, Voltage-clamp commands to another Cs^+ -loaded cell exposed to $1 \mu\text{g ml}^{-1}$ TTX. Frames 1 and 2 show hyperpolarizing commands and the small leakage current. Frames 3 and 4 show depolarizing commands that activate an inward-going current that shows little or no decay during the command. (See *b* for value of holding current.) *b*, I - V curve before and after applying a drop of 10 mM Co^{2+} . The region of negative slope conductance is absent after Co^{2+} . Although a more negative holding potential was used after Co^{2+} , the depolarizing commands used reach well into the previous region of negative slope conductance, and no inward-going current is observed. This impairment was lost before measurements at the previous holding potential could be made. *c*, Voltage-clamp commands after application of Co^{2+} . The inward-going current is no longer present during the depolarizing commands.

depolarization, and the I - V curve shown in Fig. 3b displays no region of negative slope conductance. These data suggest that a slow, voltage-dependent, inward current is present in these cells, probably carried by Ca^{2+} ions. However, a TTX-resistant Na^{+} current, which is affected by Ba^{2+} and Co^{2+} in the above manner, cannot be excluded. Although Cs^{+} injection substantially reduced the outward-current contribution to the total membrane current, it did not block it completely. During large depolarizing commands (for example >20 mV from the holding potential), an outward current, possibly a Ca^{2+} -dependent K^{+} current, developed. This latter current apparently was less sensitive to Cs^{+} injection and contaminated our inward current measurements.

We appreciate the difficulties associated with attempting to voltage clamp neurones whose geometry is as complex as that of the hippocampal pyramids. However, an analysis of the cable properties of CA3 neurones suggests that their total electrotonic length is less than one space constant¹¹. A twofold increase in input resistance as a result of Cs^{+} injection theoretically would reduce the electrotonic length by a factor of $1/(2^{1/2})$, improving spatial control of voltage. Given such an electronic structure and twice the average membrane time constant of 23.6 ms (ref. 11), equation (32) of Rall¹⁷ predicts a time constant of decay of about 7 ms for the passive current transient following a step change in potential at the soma. Empirically, we have found that this decay is complete in less than 7 ms.

After Cs^{+} injection, the persistent inward current exerts almost complete control over the membrane potential in unclamped conditions. Small applied currents move the membrane potential rapidly into and out of the region of negative slope resistance (approximately -50 to -10 mV). This sensitivity to small potential changes suggests that the slow inward current is generated electrically near the recording site and is therefore amenable to voltage-clamp study.

When not voltage clamped, all of the neurones illustrated displayed spontaneous bursting behaviour after Ba^{2+} application or Cs^{+} injection, even in the presence of TTX. The bursts consisted of one or more regenerative events, presumably Ca^{2+} spikes, triggered by slow depolarizations of the membrane potential. The voltage-clamp data and the behaviour of the unclamped cell are therefore in agreement.

Although the SEC is not capable of resolving currents underlying the action potential, we believe that it is more than adequate for resolution of the currents described here. This slow inward current in CA3 neurones is similar to the slowly inactivating inward current shown to be responsible for endogenous bursting in molluscan and spinal neurones³⁻⁶, which suggests that the burst tendencies of CA3 neurones may arise from similar mechanisms. The results of the experiments with Cs^{+} injection indicate that K^{+} currents may be important in modifying the effect of an inward Ca^{2+} current. As in molluscan neurones¹⁸, several K^{+} currents undoubtedly exist and may control different aspects of bursting.

This work was supported by grants NS11535 and NS15772 from the National Institute of Neurological and Communicative Disorders and Stroke, NIH, USPHS.

Received 18 February; accepted 2 April 1980.

- Kandel, E. R. & Spencer, W. A. *J. Neurophysiol.* **24**, 243-259 (1961).
- Wong, R. K. S. & Prince, D. A. *Brain Res.* **159**, 385-390 (1978).
- Wilson, W. A. & Wachtel, H. *Science* **186**, 932-934 (1974).
- Eckert, R. & Lux, H. D. *J. Physiol., Lond.* **254**, 129-152 (1976).
- Johnston, D. *Brain Res.* **107**, 418-423 (1976).
- Partridge, L. D., Thompson, S. H., Smith, S. J. & Connor, J. A. *Brain Res.* **164**, 69-79 (1979).
- Schwindt, P. C. & Crill, W. E. *Brain Res.* **120**, 173-178 (1977).
- Wilson, W. A. & Goldner, M. M. *J. Neurobiol.* **6**, 411-422 (1975).
- Schwartzkroin, P. A. *Brain Res.* **85**, 423-436 (1975).
- Schwartzkroin, P. A. *Brain Res.* **128**, 53-68 (1977).
- Johnston, D. *Biophys. J.* **25**, 304a (1979).
- Sperelakis, N., Schneider, M. F. & Harris, E. J. *J. gen. Physiol.* **50**, 1565-1583 (1967).
- Magura, I. S. *J. Membrane Biol.* **35**, 239-256 (1977).
- Hagiwara, S. in *Membranes—A Series of Advances* Vol. 3 (ed. Eisenman, G.) 359-381 (Dekker, New York, 1975).
- Akaike, N., Lee, K. S. & Brown, A. M. *J. gen. Physiol.* **71**, 509-531 (1978).
- Tillotson, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1497-1500 (1979).
- Rall, W. *Biophys. J.* **9**, 1483-1508 (1969).
- Thompson, S. H. *J. Physiol., Lond.* **265**, 465-488 (1977).

Circadian clock in *Limulus* brain increases response and decreases noise of retinal photoreceptors

E. Kaplan* & R. B. Barlow Jr†

Marine Biological Laboratory, Woods Hole, Massachusetts 02543

The sensitivity and structure of ommatidia in the *Limulus* lateral eye exhibit circadian rhythms¹⁻³. The cyclic changes are mediated by the activity of efferent optic nerve fibres¹ that originate in the brain and terminate in the retina^{4,5}. We report here that the effects of the efferent input to the retina are detectable in single photoreceptor cells. At night the efferent input decreases photoreceptor noise (discrete waves of the membrane potential in darkness) and increases the photoreceptor response to light (amplitude of the receptor potential).

The receptor potential is a fundamental component in the process of visual excitation. It is believed to be formed in *Limulus* photoreceptors by the summation of discrete waves of membrane potential⁶. The statistical behaviour of the waves in light is consistent with the notion that they are triggered by single photon absorptions^{7,8}. Discrete waves occurring in the dark are thought to result from thermal events in the excitatory process^{9,10}. We found that the efferent input to the retina at night reduces the frequency of discrete waves occurring in the dark without reducing the efficacy of light to initiate them. Apparently the circadian clock can affect early steps in the process of visual excitation. The action of the clock suggests the existence of separate pathways for the spontaneous and light-evoked excitation of photoreceptor cells.

Figure 1 shows records of the membrane potential of a single, dark-adapted photoreceptor (reticular cell) *in situ*. On the day of this experiment an animal was taken from a natural light-dark environment in the early afternoon and fastened to a rigid platform in a seawater aquarium located in a light-proof cage. We removed a small section of cornea from one lateral eye and inserted a glass microelectrode into the exposed patch of retina. We impaled a single reticular cell in the late afternoon and optically isolated the ommatidium containing the cell with a light pipe aligned in contact with the cornea¹¹. The cage was then closed to dark adapt the animal. Details of the technique have been described elsewhere¹². The membrane potential of the reticular cell was recorded continuously from 6 to 10 p.m. We chose this particular period because the results of previous experiments indicated that efferent fibres in the optic nerve trunk become active about the time of sunset (~ 7 p.m.) and exert their maximal effects on the retina in the late evening hours (~ 10 p.m.)¹. Each trace in Fig. 1 shows the membrane potential recorded from the cell in darkness for 20 s and in response to a 5-s flash of constant light intensity.

At 6 p.m. numerous discrete waves of membrane potential occurred in the dark, and the response to the test flash was small. As the evening progressed, the discrete waves occurring in the dark decreased in frequency and those in the light increased. At 10 p.m. no waves occurred in the dark, whereas numerous ones summated to produce the light response. It seems that at night the photoreceptor 'noise' decreased and response increased.

Figure 2 shows that pulses of current delivered to the optic nerve during the day mimic the effects recorded from photoreceptor cells at night. The intracellular data in Fig. 2 were recorded from a single dark-adapted reticular cell. For an additional measure of photoreceptor sensitivity we also recorded the electroretinogram (ERG) from the eye using a surface

Permanent addresses: *Rockefeller University, New York, New York 10021. †Institute for Sensory Research, Syracuse University, Syracuse, New York 13210.

corneal electrode¹. The intracellular recording at 2.30 p.m. is characteristic of the day-time state of the retina, that is, it contains many discrete waves in the dark and a relatively small response to light. Note that these features are also contained in the 6 p.m. record of Fig. 1. At 2.32 p.m. periodic episodes of current pulses were delivered to the distal end of the cut optic nerve to simulate endogenous efferent activity¹. The effect of the nerve shock was to reduce the frequency of discrete waves in the dark and elevate the amplitude of both the receptor potential and ERG. Maximal response amplitudes were recorded after about 2 h of nerve shock (4.30 p.m.) at which time the current pulses were turned off. Subsequently, the receptor potential and ERG decreased in amplitude and discrete waves reappeared in the dark.

In Fig. 2 simulation of efferent optic nerve activity during the day produced night-time characteristics in the photoreceptor recordings; compare the 4.30 p.m. record with the 10 p.m. record in Fig. 1. Cessation of current pulses to the cut optic nerve at night led to the return of the day-time characteristics; compare the 8.30 p.m. record with the 6 p.m. record in Fig. 1.

The results reported here for single photoreceptor cells parallel those recorded from single optic nerve fibres *in situ*¹. In both cases the time of day when the light-evoked response is elevated and spontaneous activity is depressed corresponds to the period of endogenous efferent activity. Also, in both cases, cutting the optic nerve abolished the circadian changes. Finally,

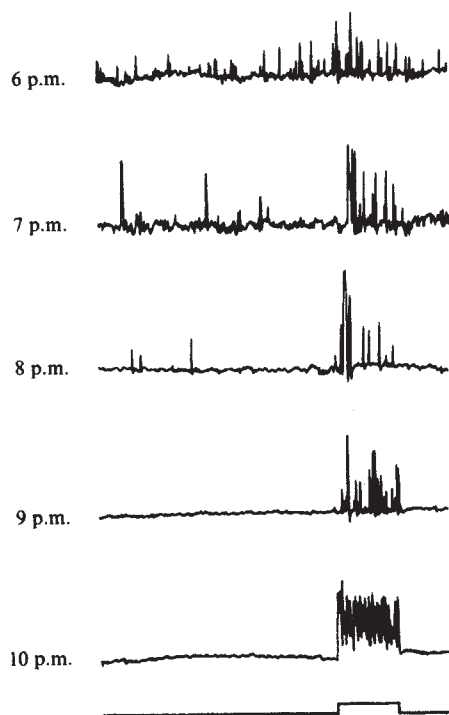


Fig. 1 Intracellular records from a retinal cell of the *Limulus* lateral eye *in situ*. The cell was impaled in the late afternoon with a glass microelectrode without disrupting the blood supply to the eye. At 5 p.m. the cage was closed to dark adapt the animal. The only light stimuli reaching the animal during the next 5 h (5 to 10 p.m.) were the 5-s test flashes (bottom trace) delivered by a light pipe isolated on the ommatidium containing the recorded retinal cell. In each trace a 20-s recording of the cell membrane potential in the dark precedes the response to the test flash. Based on the results of previous experiments, we estimate that the transmission of efferent optic nerve impulses to the retina from the circadian clock in the brain began between 6 and 7 p.m. By 9 p.m. efferent input seems to have abolished the occurrence of discrete waves in the dark and increased the response to the test flash. The recording was lost shortly after the 10 p.m. test. Cell resting potential was 50 mV. Scale bar, 10 mV. The intracellular recordings for this study were carried out with freshly collected crabs at the Marine Biological Laboratory, Woods Hole, Massachusetts.

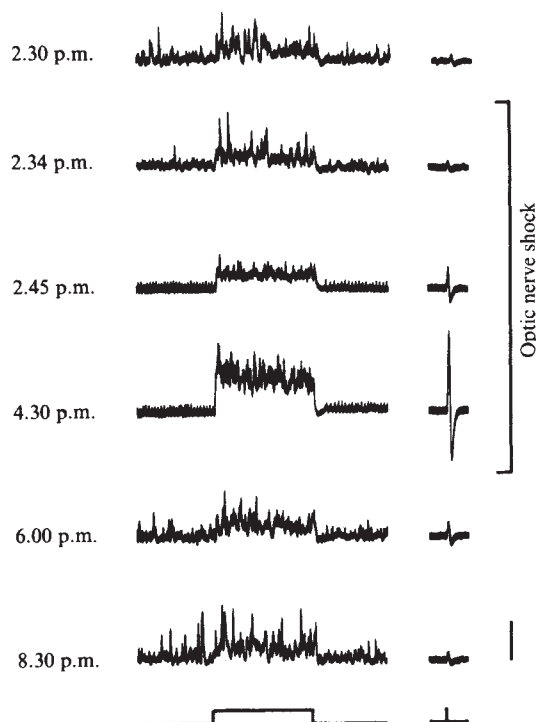


Fig. 2 Effect of optic nerve shock on visual responses of the *Limulus* lateral eye. On the left are intracellular records of a dark-adapted photoreceptor cell recorded with a glass microelectrode in response to 4-s flashes of constant intensity. On the right are electroretinograms (ERGs) recorded with a corneal electrode in response to 100-ms flashes. Before impaling the retinal cell, the optic nerve was exposed, sectioned and its distal end placed in a suction electrode. The 2.30 p.m. records are characteristic of the dark-adapted day-time state of the retina: small responses to light and frequent occurrence of discrete waves in the dark. At 2.32 p.m., current pulses were delivered to the cut nerve at the rate of 4 s^{-1} for 30 s every minute. This shock regime was chosen to simulate the maximum level of endogenous efferent activity recorded at night from the proximal end of cut optic nerves in other preparations. The effect of nerve shock was to elevate the responses to light and reduce the 'noise' in the dark. The increased response amplitudes correspond to about a 20-fold increase in visual sensitivity. The small spikes occurring in the dark are believed to originate in the second-order eccentric cells. They are often observed in experiments in which efferent activity is simulated by nerve shock but not when it is generated by the circadian clock as in Fig. 1. After the optic nerve shock was turned off at 4.30 p.m., the characteristics of the recorded responses returned to the day-time state and remained there until the cell was lost at 8.45 p.m. Cutting the optic nerve blocked the transmission of endogenous efferent activity to the retina at night and thus eliminated circadian changes of the type shown in Fig. 1. Duration of each intracellular record is 10 s. Scale bar, 10 mV for the intracellular data and 50 μV for the ERG records.

shocking the cut optic nerve during the day produced the physiological effects characteristic of the night-time state of the retina. We therefore conclude that the reduction in noise and increase in response of the photoreceptor in Fig. 1 were mediated by the endogenous activity of efferent optic nerve fibres.

The night-time increase in the amplitude of the photoreceptor response in Fig. 1 can be largely attributed to an increase in the number of photons caught by the retinal cells. Circadian changes in the morphology of the retinal cells and surrounding pigment cells maximally expose the photosensitive rhabdom to incident light collected by the corneal lenses³.

The night-time reduction in photoreceptor noise cannot be attributed to any known cellular mechanism. The spontaneously occurring discrete waves in retinal cells seem to result from the thermal isomerization of single rhodopsin molecules, whereas the light-evoked waves result from the photo-isomerization of

rhodopsin^{9,10}. This seems to be a general property of photoreceptors^{13,14}. If the spontaneous fluctuations in membrane potential are in fact initiated by thermal events, by what mechanism does the circadian input to the retina abolish the effects of thermal energy without influencing those initiated by light? Perhaps part of the chain of events triggered by light is separate from that initiated by thermal energy, and only the latter part is influenced by the efferent input to the retina. Whatever the answer, neural signals from the circadian clock seem to act at a stage preceding the excitatory conductance change in the photoreceptor membrane.

Our intracellular recordings did not reveal any additional information on the nature of the mechanisms that mediate the effects of the circadian clock. For example, no consistent changes in either the potential or the resistance of the photoreceptor membrane were detected. Long-term recordings of the type shown in Figs 1 and 2 are generally associated with a slight but steady increase in resistance as monitored by current pulses passed through the recording electrode. Such drifts in resistance

are independent of the time of day and probably result from a gradual clogging of the microelectrode. Neither membrane potential nor resistance reflect the activity of efferent optic nerve fibres.

The elevated night-time sensitivity of the retina may partially compensate for diurnal changes in ambient illumination. The circadian rhythm of sensitivity corresponds reasonably well to that of the animal's locomotor behaviour¹⁵. Interestingly, the period of elevated sensitivity also corresponds to the time the animals congregate in shallow water to mate (C. M. Cavanaugh, unpublished observations).

In sum, the efferent input from a circadian clock enhances retinal sensitivity at night by increasing the quantum catch and decreasing the spontaneous activity of the reticular cells. The efferent input thereby increases the signal-to-noise ratio of the photoreceptor response. The efferent input seems to act at an early stage in the process of visual excitation.

This research was supported by NIH grants EY-00667, EY-00108 and EY-00188 and NSF grant BNS77-19436.

Received 17 January; accepted 7 May 1980.

1. Barlow, R. B. Jr, Bolanowski, S. J. Jr & Brachman, M. L. *Science* **197**, 86-89 (1977).
2. Chamberlain, S. C. & Barlow, R. B. Jr *Biol. Bull.* **153**, 418-419 (1977).
3. Barlow, R. B. Jr & Chamberlain, S. C. in *Effects of Constant Light* (eds Williams, T. P. & Baker, B. N.) (Raven, New York, 1980).
4. Chamberlain, S. C. thesis and Spec. Rep. ISR-S-17 (Institute for Sensory Research, Syracuse University, 1978).
5. Fahrenbach, W. H. Z. *Zellforsch.* **144**, 153-166 (1973).
6. Dodge, F. A. Jr, Knight, B. W. & Toyoda, J. *Science* **160**, 88-90 (1968).

7. Yeandle, S. *Am. J. Ophthalmol.* **46**, pt 2, 82-87 (1958).
8. Adolph, A. R. *J. gen. Physiol.* **48**, 297-322 (1964).
9. Adolph, A. R. *J. gen. Physiol.* **52**, 584-599 (1968).
10. Srebro, R. & Behbehani, M. *J. Physiol., Lond.* **224**, 349-361 (1972).
11. Barlow, R. B. Jr *J. gen. Physiol.* **54**, 383-396 (1969).
12. Barlow, R. B. Jr & Kaplan, E. *J. gen. Physiol.* **69**, 203-220 (1977).
13. Yau, K. W., Lamb, T. D. & Baylor, D. A. *Nature* **269**, 78-80 (1977).
14. Yau, K. W., Matthews, G. & Baylor, D. A. *Nature* **279**, 785-786 (1979).
15. Barlow, R. B. Jr & Palfai, T. in *Research Report LSC-14 4* (Institute for Sensory Research, Syracuse University 1971).

Light-induced calcium fluxes from outer segment layer of vertebrate retinas

S. Yoshikami, J. S. George & W. A. Hagins

Laboratory of Chemical Physics, NIAMDD, National Institutes of Health, Bethesda, Maryland 20205

Light affects the membrane potential of a vertebrate retinal rod or cone cell by transiently reducing its 'dark current', an inward flow of Na⁺ through the plasma membrane covering the light-absorbing outer segment^{1,2}. In rods most photons are absorbed by photopigment molecules in intracellular disk membranes that are physically separate from the plasma membrane²⁻⁵. Photochemical events in the disks probably control the Na⁺ conductance of the plasma membrane through a chain of events that ultimately change the level of a diffusible substance in the cytoplasm of the outer segment. By combining reversibly with Na⁺ carriers or channels in the plasma membrane, this 'internal transmitter' substance would modulate the dark current and thus the rod's membrane potential⁶. There is evidence that the internal transmitter is free Ca²⁺ (refs 7-10). Ca²⁺ buffer EGTA has a desensitizing effect on rods, consistent with a free cytoplasmic Ca²⁺ concentration of about 1 μM in darkness and release of 500-1,000 additional calcium ions for each absorbed photon¹⁰. However, there is insufficient proof that Ca²⁺ is released into the cytoplasm of rod outer segments by light and removed again in darkness. Reports that light reduces the total Ca²⁺ in outer segments have varied widely in their details¹¹⁻¹⁵. Small amounts of Ca²⁺, ~1 mol per mol of photoisomerized rhodopsin, are released by strongly illuminated isolated rod disks¹⁶, but this is not enough to explain the large size of single photon responses^{10,17}. We have studied net movements of Ca²⁺ in the receptor layer of albino rat retinas during the responses of the rods to light with the aim of detecting release of internal Ca²⁺ by a rise in its extrusion into the interstitial space.

Microelectrodes sensitive to free Ca²⁺ were made as described previously¹⁸. The siliconed tips were 2-3 μm in diameter and were filled for a length of 250 μm with *O*-nitrophenyl octyl ether containing 10% Na tetraphenylborate and 10% of the Ca²⁺ selective synthetic ionophore, ETH 1001 (Orion)¹⁹. The remainder of the electrode contained 0.1 M CaCl₂. Electrode resistances were 1-3,000 MΩ. The microelectrodes responded to step changes in Ca²⁺ activity from 0.5 to 5 mM with potential changes having an exponential rise of 130 ± 10 ms. Over the range of free [Ca²⁺], *a*_{Ca}, from 10⁻⁷ to 10⁻² M, the Ca electrode potential *V*_{Ca} followed the equation

$$V_{Ca} = V_o + M \log_{10}(a_{Ca} + a_o) \quad (1)$$

where *V*_o is a constant dependent on the solutions in the Ca and reference electrodes and *M* is 31 ± 1 mV per decade change in *a*_{Ca} at 37 °C. The apparent basal Ca activity due to electrode permeability to other ions, *a*_o < 5 × 10⁻⁸ M. Changes in pH of 1.5 unit on a background of 70 μM Ca²⁺ produced electrode potential changes of <20 μV. The electrodes were insensitive to changes in Mg²⁺, Na⁺, and K⁺ (ref. 19).

The Ca electrode potential was referred to one of a pair of Ringer-filled micropipettes mounted in a cluster with it (shown in Fig. 1a). The entire assembly could be thrust under direct vision by IR microscopy¹ into the upturned receptor surface of an isolated retina so that the external voltage gradient *V*_d due to the dark current could be monitored with the change in potential of the Ca²⁺ electrode. As shown in Fig. 1a, electrode B and the Ca electrode were placed at the same depth in the retina so that the Ca signal was not mixed with light-induced changes in *V*_d.

Isolated rat retinas were prepared as previously described^{1,2}. They were mounted in a chamber in a stream of physiological solution that passed parallel to the outer segment layer in laminar flow. The fluid velocity 50 μm above the rod tips was ~100 μm s⁻¹. An arrangement of valves allowed the fluid composition to be changed in 5 s. When more rapid changes in fluid composition were needed, a continuous jet of the desired fluid from a pipette in the bath was quickly lowered so as to sweep over the retina near the electrodes.

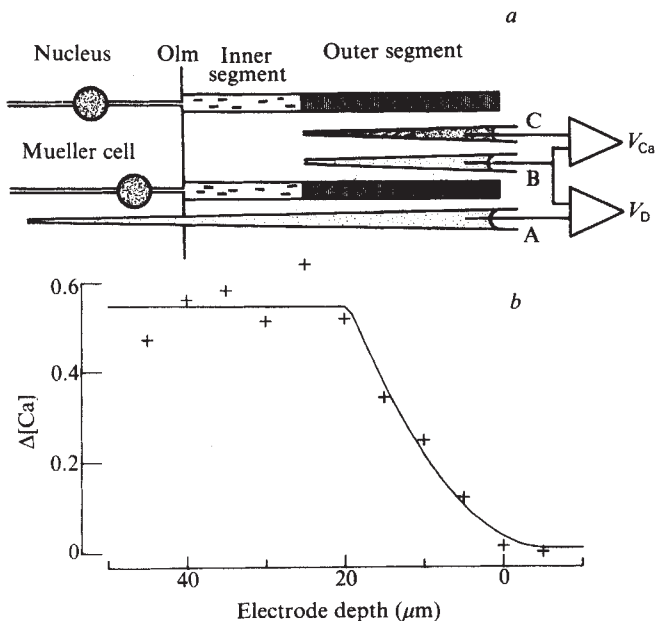


Fig. 1 Spatial distribution of the light-induced transient increase in interstitial $[Ca^{2+}]$ with depth of the Ca-sensing electrode in the receptor layer. *a*, Diagram of the structure of the receptor layer to scale indicated at the bottom of *b*. Electrodes placed as shown in the interstitial spaces between the rods. *b*, Amplitude of the Ca^{2+} against depth. Each point is the mean area under the first 800 ms of the Ca transient obtained by averaging 16 flash responses. Curve: amplitude distribution to be expected from a source of Ca^{2+} limited to the outer segment layer of the retina. The assembly of three electrodes was moved in steps through the receptor layer. Stimuli: 2 μ s, 560 nm, 110 photons absorbed per rod per flash. Ambient free $[Ca^{2+}]$ 10 μ M in a Ringers of composition similar to that of Fig. 2a.

Light causes a transient increase in the Ca^{2+} activity in the interstitial space of the outer segment layer. Figure 2a shows the effect of a 560-nm flash yielding 110 photons absorbed per rod on V_d recorded between electrodes A and B at depths of 75 and 25 μ m from the rod tips and changes in the Ca^{2+} activity seen by the Ca electrode at 25 μ m. The flash immediately shut down the dark current for about 1 s after which V_d returned to its initial level with a small overshoot. This waveform is characteristic of rat rods in the 70 μ M unbuffered Ca^{2+} of the Ringer used. The Ca electrode showed a transient increase in a_{Ca} followed by a longer lasting decrease to slightly below the level before the flash. Although V_{Ca} was logarithmic in Ca^{2+} activity (equation (1)), the changes seen were so small relative to the background a_{Ca} that the relationship between ΔV_{Ca} and Δa_{Ca} was nearly linear. The time integral of the Ca^{2+} transient (smooth line) was nearly zero over the duration of the record. The rise in Ca^{2+} was ~ 1.2 μ M at its peak just after the minimum in V_d .

If the Ringer was augmented with 10 mM Ca^{2+} buffer, nitrilotriacetic acid (NTA), and the Ca^{2+} and Mg^{2+} activities carefully adjusted to those of the control, the light-induced change in V_d was unaffected, but the Ca^{2+} transient was greatly reduced (Fig. 2b). This effect of NTA could not be attributed to its action as a buffer for H^+ or Mg^{2+} : there was 10 mM of the pH buffer HEPES present in both cases, and NTA is a weak Mg^{2+} buffer at the concentration used (500 μ M Mg^{2+}). Thus the responses of the Ca electrode were probably due to changes in interstitial a_{Ca} .

With a fixed 10- μ M free Ca^{2+} concentration in the Ringer and a stimulus of 80 photons absorbed per rod per flash, the size and shape of the Ca transient varied with electrode depth in the retina. Figure 1b shows data from Ca transients recorded at 5- μ m intervals beginning 5 μ m above the tips of the outer segments and extending inward to a depth of 45 μ m. The points were computed from the time integrals of the responses 800 ms after the flashes. This time was near the peaks of the transients themselves. Small Ca transients could usually be seen even when the electrode was 25 μ m above the retinal surface. These were much slower than the intraretinal ones and were caused by

clouds of Ca^{2+} , released from distant retinal areas, drifting past the Ca electrode. In the space between the rods, the transients grew rapidly in size with increasing electrode depth in the region between the rod tips and the inner-outer segment junction at 25 μ m depth. As the Ca electrode was advanced further, the transients remained at constant size or decreased slightly. In the dark, there was no detectable gradient of a_{Ca} in the extracellular space. Thus the Ca transients must indicate increased efflux and not a light-induced fall in a steady dark influx.

Viewed as a problem in one-dimensional diffusion of Ca^{2+} from sources in the receptor layer to sinks there and in the flowing Ringer stream, the concentration c of Ca^{2+} should follow the differential equation

$$\frac{\partial c}{\partial t} - D \frac{\partial^2 c}{\partial x^2} = F(x, t) \quad (2)$$

where D is the diffusion coefficient of Ca^{2+} in the interstitial space perpendicular to the retinal surface and $F(x, t)$ is the distribution of Ca^{2+} sources and sinks on the receptor cells. In

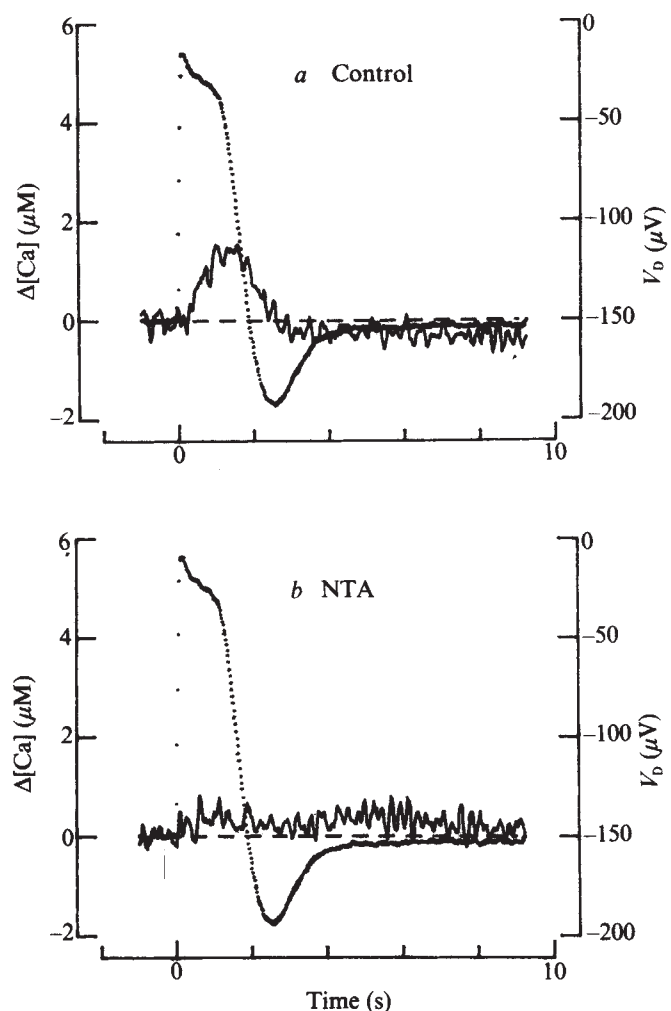


Fig. 2 Changes in interstitial dark voltage gradient V_d and free $[Ca^{2+}]$ in the rod layer of an albino rat retina. Voltage-recording and Ca-recording electrodes placed as shown in Fig. 1a. Stimuli: 2 μ s flashes of wavelength 540–580 nm that resulted in 110 photons absorbed per rod per flash. Dotted curves: changes in V_d . Continuous noisy curves: interstitial free $[Ca^{2+}]$. Dashes: drawn at signal levels seen just before flashes. Each curve is the mean of 16 responses. *a*, Control Ringer containing 136 mM NaCl, 0.5 mM $MgCl_2$, 2.7 mM KCl, 0.5 mM NaH_2PO_4 , 11 mM glucose, 10 mM HEPES, and $CaCl_2$ added until the Ca electrode indicated a free $[Ca^{2+}]$ of 70 μ M. pH 7.14. *b*, Bathing Ringers same as control except that 10 mM nitrilotriacetic acid (NTA) was added and excess Ca^{2+} and Mg^{2+} were added by titration to achieve the same free metal concentrations as in *a*. The buffering effect of the NTA was negligible for H^+ and Mg^{2+} but reduced the concentration elevating effect of added Ca^{2+} to <10% of that of the control Ringers.

principle the two partial derivatives could be measured directly with electrode arrays and F computed as Ca^{2+} current sources and sinks¹, but the noise in the present data makes this procedure inaccurate. However, if the Ca transients are sampled at their peaks as in Fig. 1b, $\partial c/\partial t$ is negligible and equation (2) becomes an ordinary differential equation linking $c(x)$ with the Ca^{2+} source distribution $F(x)$. In Fig. 1 $c(x)$ is computed from equation (2) for a hypothetical distribution $F(x)$ (solid line) where the source is a uniform zone covering the entire length of the outer segments. No flow of Ca^{2+} is assumed to occur at the outer limiting membrane at $x = 45 \mu\text{m}$.

The data fit the hypothetical source distribution closely when its amplitude is properly scaled. In the outer segment layer, the transients grow in size roughly as the square of the electrode depth and not linearly as would be expected for Ca sources concentrated at the inner segments or more deeply. Unlike the computed curve, however, the data suggest some loss of Ca^{2+} towards the deeper retinal layers. The slight shift of the measured Ca source distribution (Fig. 1b) relative to the retinal elements of Fig. 1a is due to the $5\text{-}\mu\text{m}$ uncertainty in electrode position. Most of the Ca^{2+} arises in the region of the rod outer segments.

The internal a_{Ca} of rat rod outer segments is $\sim 1 \mu\text{M}$ (ref. 10), and that of the external medium is much higher. Thus rods, like other eukaryotic cells, must regulate their internal Ca^{2+} with membrane pumps that exchange Ca^{2+} for Na^+ (refs 20, 21) or extrude Ca^{2+} via nucleotide phosphatases in the plasma membrane^{22,23}. Is the light-induced Ca^{2+} efflux caused by a rise in the cytoplasmic a_{Ca} or by a light-stimulated increase in the activity of some Ca^{2+} extrusion mechanism? The results shown in Fig. 3 suggest the first mechanism perhaps augmented by the second through light-induced rod hyperpolarization.

If the NaCl in the bath is replaced with the depolarizing salt, potassium isethionate, the large V_d response shown in Fig. 3a (right) is completely abolished, but the interstitial increase in a_{Ca} , while slower and four times smaller, persists (Fig. 3b). Replacing the Na^+ in the medium with choline ions produces a similar result even though the rod membrane potentials should be hyperpolarized to near K^+ equilibrium potential in this solution. Since the residual Na^+ in both choline and K isethionate Ringers was less than $10 \mu\text{M}$, it is unlikely that $\text{Na}:\text{K}$ exchange causes the Ca^{2+} efflux in these solutions.

Can light still produce a Ca^{2+} efflux when plasma membrane calcium pumps are partly shunted by a divalent metal ionophore? If the rods are exposed to X-537A, the dark current and the V_d responses are greatly reduced at external $a_{\text{Ca}} > 10^{-5} \text{M}$ (ref. 8), but the light-induced rise in interstitial a_{Ca} remains large (Fig. 3d). Reducing the ambient a_{Ca} to $10 \mu\text{M}$ increases the dark current and V_d , but the Ca signal decreases slightly. Since X-537A increases the permeability of the plasma membrane to Ca^{2+} , it is likely that the increased Ca^{2+} efflux follows mainly from an increase in cytoplasmic a_{Ca} . Because the external $[\text{Ca}^{2+}]$ is greater than that in the cytoplasm even in the presence of X-537A, however, the Ca extrusion machinery must link the presumed light-induced rise in internal Ca^{2+} to that seen extracellularly, making it larger and faster.

If intracellular Ca^{2+} release causes the dark current shutdown, the kinetics of Ca^{2+} efflux should closely track the V_d waveform. To compare the kinetics of the two processes, the observed waveform of the transient increase in interstitial a_{Ca} must be transformed mathematically to allow for the finite volume v of the interstitial space (1 cm^2) and for the loss of Ca^{2+} by diffusion into the flowing Ringers. Neglecting the small unstirred layer at the retinal surface and solving equation (2) for the condition of no flow at the outer limiting membrane (OLM) (Fig. 1a) and for a uniform Ca^{2+} source distributed along the outer segments that instantaneously dumps $Q \text{ mol cm}^{-2}$ into the interstitial space yields a transient increase in a_{Ca} at $25 \mu\text{m}$ depth in the receptor layer that is close to a truncated exponential of the form

$$c(25, t) = 0.5 c(25, 0) \exp(-t/\tau) \quad (3)$$

where $c(25, 0) = Q/v$ and $\tau = 4l^2/\pi^2 D$ (ref. 24, equation 3.4.5,

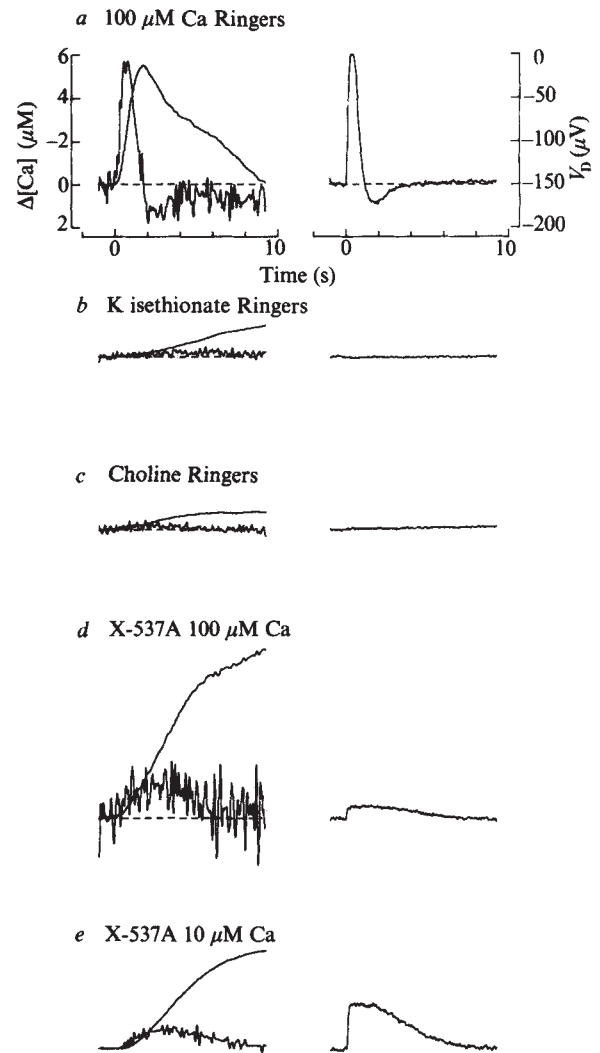


Fig. 3 Changes in dark voltage gradient V_d (right column) and extracellular free $[\text{Ca}^{2+}]$ (left column) in the rod layer of the rat retina in solutions of various compositions. Stimuli as in Fig. 2, electrode positions as in Fig. 1a. Temperature, 37°C . *a*, Control. As in Fig. 2 but Ca^{2+} added to give $p\text{Ca } 4.0$. Mean of four responses. *b*, NaCl and KCl replaced with 140 mM K isethionate, $p\text{Ca } 4.0$. Mean of 96 responses at 15-s intervals. V_d response suppressed by the depolarizing solution, but there was a distinct light-induced increase in interstitial $[\text{Ca}^{2+}]$ as shown by its time integral (smooth line in left column). The total Ca^{2+} appearing was 25% of that in the control solution. *c*, NaCl replaced with 140 mM choline chloride. This solution should hyperpolarize the rod membrane potentials but suppress the dark current. As in *b*, the V_d response disappeared but there was a measurable Ca^{2+} release from the rod layer. $p\text{Ca } 4.0$. Mean of 96 responses. *d*, $10 \mu\text{M}$ of the divalent cation ionophore, X-537A was added to control Ringers at $p\text{Ca } 4.0$. The dark current and V_d were greatly reduced, but there was a large light-induced release of Ca^{2+} from the rod layer. Mean of eight responses. The large spikes in the Ca responses were electrical interference from the bath heater. 200 photons per rod. *e*, Same retina as in *d* but ambient free $[\text{Ca}^{2+}]$ reduced to $10 \mu\text{M}$. Dark current and V_d response immediately increased, but Ca release was slightly reduced. Mean of eight responses. 200 photons per rod.

p. 101). Here l is the distance from the rod tips to the OLM. For this special position of the Ca electrode, the total outflux $J(t)$ from the entire length of outer segments thus is represented approximately by

$$2v\left(\frac{dc}{dt} + \frac{c}{\tau}\right) = J(t) \quad (4)$$

For the receptor layer of the rat retina, measurements from electron micrographs yield $v = 650 \text{ nl cm}^{-2}$. τ , Computed from a measured D for Ca^{2+} of $1.1 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ in the rod interstices (W. A. H. and S. Y., unpublished) is 650 ms . Using these two experimental parameters and forming the sum on the left-hand

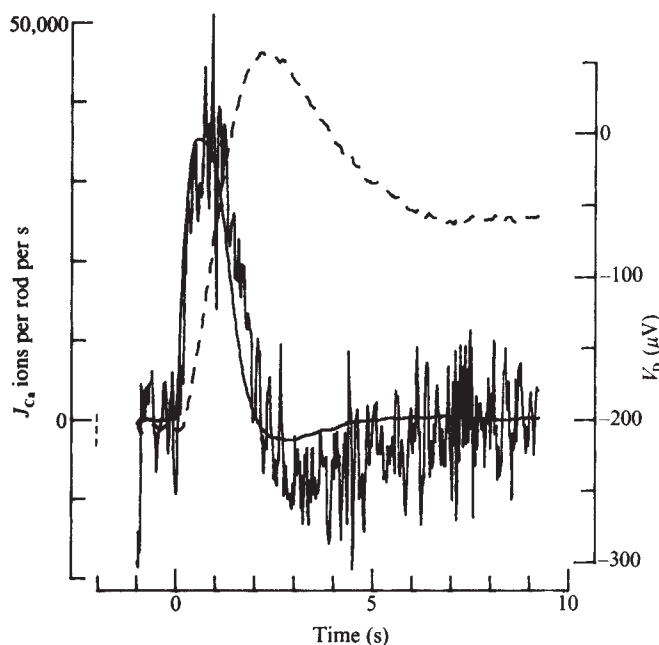


Fig. 4 Time courses of the light-induced Ca^{2+} flux from the rod layer compared with the light-induced reduction in V_a . Stimuli as in Fig. 2. Smooth line: V_a response convolved with an exponential response function of time-constant 130 ms to simulate the time-lag of the Ca electrode. Noisy curve: Ca^{2+} flux from the rods computed from equation 4. Dashes: time integral of Ca^{2+} flux in units of ions per rod on the left-hand scale. Mean of four responses to flashes of energy 110 photons absorbed per rod. The time course of the Ca efflux and the change in V_a are similar. Apparently about half of the Ca^{2+} released within 2 s of the flash is reaccumulated by the rod layer.

side of equation (4) by numerical differentiation of the experimental Ca transient yields the noisy curve in Fig. 4.

The Ca flux can now be directly compared with the waveform of the V_a response if the latter is first convolved with a truncated exponential with a 130-ms time constant to simulate the response lag of the Ca electrode. The result is a smooth line, shown in Fig. 4. The two waveforms are nearly identical, the duration of the Ca efflux being slightly longer and its peak not more than 100 ms later than the transformed V_a response.

The Ca^{2+} flux is large and biphasic. As much as half of the 45,000 ions leaving the outer segments in the first 2 s after the flash are apparently reabsorbed in the next 7 s (Fig. 4, dashed line). Since each rod absorbed 110 photons per flash, about 400 Ca^{2+} ions must have been extruded per photon response. If Ca^{2+} is the internal excitatory transmitter, the total Ca release must have been at least twice as large, because only cytoplasmic Ca^{2+} should reduce the Na^+ permeability of the plasma membrane. The observed release would only increase the cytoplasmic $[\text{Ca}^{2+}]$ by about 3 μM , a small fraction of the total Ca^{2+} in rod outer segments^{14,25}. But if a steady 25-s light stimulus of intensity 9,500 photons per rod per s is used instead of a flash, the Ca^{2+} outflux is sustained for the entire period of illumination and amounts to about 100 $\mu\text{mol l}^{-1}$ of rod water. The light-sensitive Ca store is therefore large.

The light-induced Ca^{2+} effluxes seen here are fast enough, large enough, and of the spatial distribution to be expected from the Ca hypothesis. Moreover, the observed fast exchange of Ca^{2+} across the plasma membrane of rod outer segments explains the great sensitivity of the dark current to the $[\text{Ca}^{2+}]$ in the external medium. It remains to link the Ca^{2+} fluxes to enzymatic events in the disk membranes, such as the hydrolysis of GTP²⁶⁻²⁹ and of cyclic GMP³⁰, and the binding of GDP²⁹. *Note added in proof:* Since this paper was submitted, a preliminary account has appeared in *Fedn Proc.* **39**, 2066 (1980) and similar conclusions are reached by use of surface calcium electrodes by G. H. Gold and J. Korenbrot, *Fedn Proc.* **39**, 1814 (1980).

Received 12 December 1979; accepted 12 May 1980.

- Hagins, W. A., Penn, R. D. & Yoshikami, S. *Biophys. J.* **10**, 380-412 (1970).
- Penn, R. D. & Hagins, W. A. *Biophys. J.* **12**, 1073-1094 (1972).
- Cohen, A. I. *Vision Res.* **10**, 445-453 (1970).
- Rüppel, H. & Hagins, W. A. in *Biochemistry and Physiology of Visual Pigments* (ed. Langer, H.) 257 (1973).
- Yoshikami, S., Robinson, W. E. & Hagins, W. A. *Science* **185**, 1176-1179 (1974).
- Baylor, D. A. & Fuortes, M. G. F. *J. Physiol., Lond.* **207**, 77-92 (1970).
- Yoshikami, S. & Hagins, W. A. in *Biochemistry and Physiology of Visual Pigments* (ed. Langer, H.) 245 (1973).
- Hagins, W. A. & Yoshikami, S. *Expl Eye Res.* **18**, 229-305 (1974).
- Pinto, L. H., Brown, J. E. & Coles, J. A. in *Vertebrate Photoreception* (eds Fatt, P. & Barlow, H.) 159 (1977).
- Hagins, W. A. & Yoshikami, S. in *Vertebrate Photoreceptor* (eds Fatt, P. & Barlow, H.) 97 (1977).
- Liebman, P. A. *Invest. Ophthalmol.* **13**, 700-702 (1974).
- Hendricks, T., Daeman, F. J. M. & Bonting, S. L. *Biochim. biophys. Acta* **345**, 468-473 (1974).
- Szuts, E. & Cone, R. A. *Fedn Proc.* **33**, 1471 (1974).
- Szuts, E. & Cone, R. A. *Biochim. biophys. Acta* **468**, 194-208 (1977).
- Hendricks, T., van Haard, P. M. M., Daeman, F. J. M. & Bonting, S. L. *Biochim. biophys. Acta* **467**, 175-184 (1977).
- Kaupp, U. B. & Junge, W. *FEBS Lett.* **81**, 229-232 (1977).
- Baylor, D. A., Lamb, T. D. & Yau, K.-W. *J. Physiol., Lond.* **288**, 589-611 (1979).
- Heinemann, U., Lux, H. D. & Gutnick, M. J. *Expl Brain Res.* **27**, 237-243 (1977).
- Simon, W., Ammann, D., Oehme, M. & Morf, W. E. *Ann. N. Y. Acad. Sci.* **307**, 52-70 (1978).
- Reuter, H. & Seitz, N. J. *J. Physiol., Lond.* **195**, 451-470 (1968).
- Mullins, L. J. *J. gen. Physiol.* **70**, 681-695 (1977).
- Blaustein, M. P., Ratzlaff, R. W. & Kendrick, N. K. *Ann. N. Y. Acad. Sci.* **307**, 195-212 (1978).
- Baker, P. F. *Ann. N. Y. Acad. Sci.* **307**, 250-268 (1978).
- Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn (Oxford University Press, London, 1959).
- Hagins, W. A. & Yoshikami, S. *Ann. N.Y. Acad. Sci.* **264**, 314-325 (1975).
- Robinson, W. E. & Hagins, W. A. *Biophys. J.* **16a** (1977).
- Robinson, W. E. & Hagins, W. A. *Photochem. Photobiol.* **29**, 693 (1979).
- Wheeler, G., Matsuo, Y. & Bitensky, M. W. *Nature* **269**, 822-824 (1977).
- Robinson, W. E. & Hagins, W. A. *Nature* **280**, 398-400 (1979).
- Yee, R. & Liebman, P. A. *J. Biol. Chem.* **253**, 7902-7909 (1978).

Co-infection by lactic dehydrogenase virus and C-type retrovirus elicits neurological disease

Larry R. Pease & William H. Murphy

Department of Microbiology and Immunology, The University of Michigan Medical School, Ann Arbor, Michigan 48109

Intraperitoneal injection of neuropathogenic strains of lactic dehydrogenase virus (LDV)¹⁻³ causes a histologically distinctive⁴ fatal paralytic disease characterized by an inflammatory destruction of motor neurones in the brain stem and cord in C58 mice aged over 9 months. To elicit the disease in the naturally susceptible C58 strain requires an age-associated⁵ or X-ray induced⁶ loss of immunological competence⁵, LDV infection¹⁻³ and genetic susceptibility^{1,6}. Genetic studies^{1,7} of the common inbred mouse strains showed that susceptibility to the disease was not linked to the major histocompatibility complex but correlated with the FV-1^a allele, susceptibility to spontaneous leukaemia, and infection by neuropathogenic strains of LDV¹⁻³. These observations suggested that neuropathogenic strains of LDV elicit the disease only in those strains of mice that carry multiple copies of N-tropic C-type retroviruses in their genomes^{8,9} and that are permissive for retrovirus replication. Presumably the expression of these viral genomes (high titres of virus in tissues correlating with age⁹) is the important factor. Here we present genetic evidence to support this hypothesis and briefly discuss the possible implications.

In Table 1 the mice in group 1 were selected to test whether there was a correlation between susceptibility to the disease, the number of copies of N-tropic C-type retroviruses in their genomes^{8,9}, their haplotype and their FV-1 genotype. The FV-1^b allele¹⁰ dominantly restricts N-tropic C-type retrovirus replication. All of the strains (a-e) that were FV-1^a and are known to contain multiple copies of N-tropic C-type retroviruses in their genomes^{8,9} were susceptible. The data for the congenic AKR strains are of interest because they show that

Table 1 Susceptibility of inbred mice to paralytic disease

Group	Inbred strains*	H-2 type	FV-1 type	Genomic copies of C-type retroviruses	Incidence of paralysis† age (months)		
					6	9	12
1	a C58/J	k	n	Many		10/10	15/15
	b AKR/Boy	k	n	Many		13/20	
	c AKR-H-2 ^b /Boy	b	n	Many	0/17		5/6
	d C3H/Fg Boy	k	n	Many			16/16
	e PL/J	u	n	Many			6/11
	f AKR-FV-1 ^b /Boy	k	b	Many		0/24	
	g C3H/HeJ	k	n	Few	0/8		0/12
	h DBA/2J	d	n	Few	0/13		0/14
	i B6/Boy	b	b	Few		0/13	
	j B10	b	b	Few	0/19		0/10
F ₁ hybrids	k B6-FV-1 ⁿ /Boy	b	n	Few		0/12	
	l B10.BR	k	b	Few	0/18		
	a C58/wm × C3H/HeJ	k/k	n/n		2/8	7/12	
	b C58/wm × DBA/2J	k/d	n/n		2/8	5/8	
	c C58/wm × B10	k/b	n/b		0/26		0/10
2	d C58/wm × B10.BR	k/k	n/b		0/20		0/14
	e C58/wm × BALB/cWm	k/d	n/b		0/11		0/19
Backcross progeny							
3	a (C58 × B10) × B10	k/b or b/b	n/b or b/b		0/43		
	b C58 × (C58 × B10)	k/k or k/b	n/n or n/b		18/53		6/12
	c (C58 × B10) × C58	k/k or k/b	n/n or n/b		1/93		1/20

* In group 1, strains *a*–*c* are considered to have a 'high' incidence of leukaemia ranging from 40 to 90% depending on the colony. Strains *d* and *e* have a 'moderate' incidence of leukaemia ranging from 20 to 40%. The remaining mice listed have a 'low' incidence of spontaneous leukaemia (usually < 5%)¹⁶. AKR-H-2^b/Boy is congenic to AKR/Boy and is H-2^b rather than H-2^k. AKR-FV-1^b/Boy is congenic to AKR/Boy but is FV-1^b rather than FV-1ⁿ. B6/Boy is a subline of C57BL/6 and is FV-1^b. The congenic line B6-FV-1ⁿ/Boy is FV-1ⁿ. The C58/wm and BALB/cWm mice are our stock inbred lines. The following abbreviations were used: C58 for C58/wm, B10 for C57BL/10Sn, and B10.BR for B10.BR/SgSn. In Groups 2 and 3 the female is listed first for each breeding set; mice were 2–5 months old when mated.

† Mice received 550R of whole body X-ray irradiation 24 h before the intraperitoneal injection of LDV (10⁷–10⁸ infectious doses) derived from either line *1b* cells or from a stock prepared from the plasma of infected BALB/cWm mice². The plasma stocks of LDV cause the same clinical and histopathological effects as line *1b* derived virus². After infection, mice were held for 30 days and scored for the incidence of paralysis. Paralysed and disease-free mice (3–6 per group) were killed and the diagnosis of paralysis or its absence confirmed histologically as described previously⁴. The sex of mice did not affect their susceptibility in any of the test groups.

when H-2^b was substituted for H-2^k, mice were still susceptible (compare *c* with *b*). When FV-1^b was substituted for FV-1ⁿ, mice were resistant (compare *f* with *b*). The mice in groups *g* and *h* were representative of strains that carry few copies of retroviruses in their genomes but are permissive (FV-1ⁿ) for retrovirus replication. None contracted the disease. The strains in groups *i* and *j* contain few copies of retroviruses in their genomes, are not permissive (FV-1^{b/b}) for N-tropic C-type retrovirus replication, and were not susceptible. Introduction of the FV-1ⁿ allele in place of FV-1^b (compare *k* with *i*) or H-2^k in place of H-2^b (compare *l* with *j*), did not confer susceptibility. These results confirmed our findings^{1,7} that susceptibility or resistance was not H-2 linked and indicated that mice had to carry multiple copies of N-tropic C-type retroviruses in their genomes and be permissive for retrovirus replication to be susceptible.

The mice in group 2 were used to test whether mice presumed to be genetically susceptible, but containing too few copies of N-tropic C-type retroviruses in their genomes (mice in *g* and *h*, Group 1), would manifest paralytic LDV infection if they became infected by the appropriate C-type retrovirus. Thus, 'genetically susceptible' male mice 'deficient' in virus (C3H/HeJ and DBA/2J) were bred to C58 females. The assumption was that the C58 mothers would transmit virus to the newborn F₁ progeny either chromosomally or by the milk. Table 1 shows that the F₁ hybrids derived from the 'potentially susceptible strains' (groups *a* and *b*) were susceptible. The corresponding genetically resistant (FV-1^{n/b}) hybrids were not (groups *c*–*e*) even when the H-2^k haplotype (group *d*) was substituted for H-2^b (group *c*).

In group 3, test-cross progeny were tested to determine whether resistance segregated in accordance with the hypothesis that the FV-1^b allele determined resistance. The data show that all of the backcross progeny in control group *a*, which in

principle inherited an FV-1^b allele, were resistant. In theory, 50% of the test-cross progeny in group *b* should have been FV-1^{n/n} and susceptible if they were infected by an N-tropic C-type retrovirus by their mothers. The results confirmed this expectation with a somewhat lower incidence of disease (34%) in the incompletely susceptible 6-month-old indicator C58 mice^{1,6,7}. Moreover, when backcross progeny were nursed on resistant (C58 × B10) mothers (group *c*) there was a strong maternal effect. Such an effect would be consistent with the failure to transmit virus to the progeny by the milk¹¹, or the transmission of antibody to the progeny by the milk, thereby protecting them from C-type retrovirus infection. Other mechanisms of conferring maternal resistance are not ruled out by the data¹¹.

If one accepts the genetic data at their face value they support the hypothesis that dual infection by LDV and N-tropic C-type retrovirus(es) was required to elicit a specific paralytic motor neurone disease in genetically susceptible mice. Expression of N-tropic C-type retrovirus genomes in FV-1^{n/n} mice seemed to be an important factor in the susceptible strains. The paradox is that neither LDV¹² or the common C-type retroviruses are frankly cytopathic to cells either *in vitro* or *in vivo*. Although Gardner *et al.*¹³ and McCarter *et al.*¹⁴ have reported the occurrence of neuropathogenic C-type viruses in mice, LDV has not been evaluated as an essential contributory factor in such virus stocks. However, these investigators also have not claimed that neurone destruction is a direct cytopathic response to viral infection. The striking feature of the model we describe is that two viruses that are ubiquitous in mice, and that ordinarily do not cause overt disease, seem to act together to elicit a specific disease syndrome in a genetically susceptible host. These findings are of interest because they direct attention to how unrelated viruses may interact together at the cellular level, provide an additional way to characterize phenotypically LDV

and mouse leukaemia virus strains, and may provide an additional insight to the complex interactions between viruses and their hosts in the slow viral infections of the nervous system¹⁵.

This work was supported by funds from the Institute of Gerontology of the University of Michigan and by USPHS grant S07-RR-05383. All the 'Boy' mouse strains were from Dr Edward Boyse. The mice with a J designation and B10 and B10.BR lines were from Jackson Memorial Laboratory.

Received 18 February; accepted 27 May 1980.

- Murphy, W. H., Nawrocki, J. F. & Pease, L. R. in *Animal Models of Neurologic Disease* (eds Behan, P. O. & Rose, C. F.) (Pitman, London, in the press).
- Nawrocki, J. F., Pease, L. R. & Murphy, W. H. *Virology* **103**, 259–264 (1980).
- Martinez, D., Brinton, M. A., Tachovsky, T. G. & Phelps, A. H. *Infect. Immun.* **27**, 979–987 (1980).
- Lawton, J. W. M. & Murphy, W. H. *Archs Neurol.* **28**, 367–370 (1973).
- Murphy, W. H., Tam, M. R., Lanzi, R. L., Abell, M. R. & Kauffman, C. *Cancer Res.* **30**, 1612–1622 (1970).
- Duffey, P. S., Martinez, D., Abrams, G. D. & Murphy, W. H. *J. Immun.* **116**, 475–481 (1976).
- Pease, L. R., Abrams, G. D. & Murphy, W. H. *J. Immun.* (submitted).
- Lowy, D. R., Chattopadhyay, S. K., Teich, N. M., Rowe, W. P. & Levine, A. S. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3555–3559 (1974).
- Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. & Rowe, W. P. *Cold. Spring Harb. Symp. quant. Biol.* **39**, 1085–1101 (1974).
- Lilly, F., Duran-Reynals, M. L. & Rowe, W. P. *J. exp. Med.* **141**, 882–889 (1975).
- Melief, C. J. M., Louie, S. & Schwartz, R. S. *J. natn. Cancer Inst.* **55**, 691–698 (1975).
- Rowson, K. E. K. & Mahy, B. W. J. *Virol. Monogr.* **13**, 57–87 (1975).
- Gardner, M. R. *et al.* in *Amotrophic Lateral Sclerosis* (eds Adams, J. M., Johnson, R. T. & Brazier, M. A. B.) 217–234 (1976).
- McCarter, J. A., Ball, J. K. & Frei, J. V. *J. natn. Cancer Inst.* **59**, 179–183 (1977).
- TerMeulen, V. & Hall, W. W. *J. gen. Virol.* **41**, 1–25 (1978).
- Staats, J. *Cancer Res.* **36**, 4333–4377 (1976).

S100 protein is present in cultured human malignant melanomas

Richard Gaynor, Reiko Irie, Donald Morton & Harvey R. Herschman

Division of Hematology-Oncology, Department of Medicine; Division of Oncology, Department of Surgery; Department of Biological Chemistry and Laboratory of Nuclear Medicine and Radiation Biology, Center for the Health Sciences; University of California, Los Angeles 90024

S100 protein, so called because of its solubility in 100% saturated ammonium sulphate¹, is an acidic cytoplasmic protein specific for the nervous system^{2,3}. It is localized primarily to glial elements of the brain, Schwann cells in the peripheral nervous system and satellite cells in sympathetic ganglia^{4,5}. This brain-specific protein shows a close immunological relationship among a wide variety of vertebrates, as measured by a quantitative complement fixation technique⁶. Although strict serological conservation and tissue localization have been maintained among different species, no function for S100 protein has been determined. However, the appearance of S100 protein in brain is correlated with maturation of the nervous system in both rat and man^{7,8}. Previous *in vitro* studies on the production and regulation of S100 protein in the C6 cell line derived from a rat astrocytoma indicated that cell contact and cessation of division induced S100 protein⁹. Neuroblastoma cell lines lack this protein^{10,11} although it is found in other astrocyte cell lines. We are unaware of any reports of cell lines, other than those of glial origin, which produce S100 protein. Melanocytes are derived from neuroectodermal elements, and this common origin of glial cells and melanocytes prompted us to study whether S100 protein might be present in cell lines from human malignant melanomas. We now report the presence of S100 protein in five of seven continuous cell lines of human malignant melanoma.

All lines used in our studies were derived from metastatic melanomas, and were maintained in culture for 12–20 passages. Most lines have been characterized previously^{12,13}. After the melanoma cells in tissue culture flasks were confluent, the cells were collected and homogenized. The soluble protein was

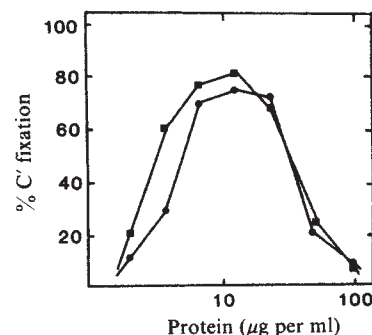


Fig. 1 Complement (C') fixation curves of S100 for human brain medulla (●) and M29 cell line (■). M29 cell line extract was prepared as described in Table 1 legend. The brain sample was homogenized with Tris buffer, 2 ml per g tissue, and spun at 40,000g for 10 min before S100 determination. Quantitative complement fixation using an 8,000-fold dilution of antiserum was carried out on extracts of M29 and human brain medulla.

assayed by quantitative complement fixation for S100 protein, using rabbit antisera to bovine S100. Complement fixation curves for extracts of human brain and one of the melanoma cell lines are shown in Fig. 1. Five of seven cell lines from human melanomas were found to have S100 protein (Table 1). In two of the lines (lines 19 and 12) the amount of S100 protein exceeded that present in human brain medulla. Lymphocyte cell lines of the two patients from which M14 and M29 were derived were negative for S100, as were HeLa cells and a human sarcoma cell line. Cells from line M14 were injected into a nude mouse; the tumour resulting after 30 days had 2.2 µg of S100 per g of tumour. Of the lines studied, only M12 and M19 were melano-

Table 1 S100 protein values for melanoma cell lines and other tissues

Specimen	µg S100 per mg soluble protein
Melanoma cell lines	
M12	6.2
M14	1.0
M15	1.9
M19	30.0
M21	ND
M24	ND
M29	3.5
Other cells and tissues	
C6 rat glial cells	2.6
Brain cerebrum (human)	2.3
Brain medulla (human)	5.6
HeLa cells	ND
Sarcoma cell line (human)	ND
Lymphocyte cell lines	
L14	ND
L29	ND

Cell lines were cultured in 75-cm² culture flasks using RPMI 1640 medium with 20% fetal calf serum and maintained in 5% CO₂. Confluent cells were washed twice with phosphate-buffered saline (NaCl, 8 g l⁻¹; KCl, 0.2 g l⁻¹; MgCl₂·6H₂O, 0.1 g l⁻¹; Na₂HPO₄·7H₂O, 2.16 g l⁻¹) and once with complement fixation buffer (0.0 M tris base, 0.14 M NaCl, 0.5 mM MgSO₄, 0.15 mM CaCl₂, 0.1% bovine serum albumin). Cells were collected by scraping into complement fixation buffer, homogenized and spun at 40,000g for 10 min. Quantitative complement fixation was carried out using a 6,000-fold dilution of antiserum as described by Levine¹⁷. S100 was quantified as described by Zuckerman *et al.*¹⁸. The serological specificity of our antiserum has previously been demonstrated by double diffusion in agar and by quantitative complement fixation analysis using soluble protein extracts of a variety of tissues as antigens^{17,18}. Purified bovine S100 was used as a standard to quantify the amount of S100 present in the soluble portion of each specimen. ND, Not detectable.

tic. Line M19 cells are the source with the greatest specific activity of S100 protein so far reported.

The serological cross-reactivity of S100 protein is strongly conserved in different species⁶. To show that the S100 protein present in human brain was immunologically identical to that present in human melanoma, we examined the serological cross-reactivity, as measured by complement fixation. The ratio of antisera required to give the same percentage of complement fixation at equivalence for heterologous compared with homologous antigens provides a quantitative measure of cross-reactivity defined as the 'index of dissimilarity' (refs 14, 15). The maximal complement fixation values for brain extracts and extracts from melanoma cell lines at three dilutions of antiserum were plotted against the log of the antiserum concentration (Fig. 2). The superimposable lines for the brain and melanoma samples indicate immunological identity for the S100 protein from these two sources (that is, an 'index of dissimilarity' of 1.0).

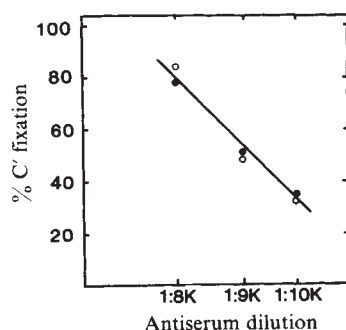


Fig. 2 Cross-reactivity of brain and melanoma cell line S100 protein. Maximal complement (C') fixation values for brain extract (●) and melanoma (○) line M29 extract at these dilutions of antisera [1/8,000 (1:8 K), 1/9,000 (1:9 K), 1/10,000 (1:10 K)] are plotted against the log of the antiserum concentration. The superimposable lines for both samples indicate immunological identity for the S100 protein from these two sources.

We do not know whether S100 protein accumulation is a function only of malignant melanocytes or of all melanocytes. S100 protein does not accumulate in human skin (our unpublished observations); however, this may reflect the low density of melanocytes normally present. Malignant melanoma cell lines have previously been shown to possess receptors for nerve growth factor¹⁶. Either nerve growth factor receptors or S100 protein may serve as a tumour-associated marker for malignant melanoma.

R.G. is a Hematology-Oncology fellow (training grant CA-09247); R.I. holds a US-PHS RCDA (CA-00543). These studies were supported by grants CA 12582 (D.M.) and DOE DE AM03 76 SF00012 (H.R.H.).

Received 28 January; accepted 14 May 1980.

- Moore, B. W. *Biochem. biophys. Res. Commun.* **6**, 739-744 (1976).
- Moore, B. W. & MacGregor, D. J. *biol. Chem.* **240**, 1647-1659 (1965).
- Moore, B. W., Perez, V. J. & Gehring, M. J. *Neurochem.* **15**, 265-272 (1968).
- Ludwin, S. K., Kosek, J. C. & Eng, L. F. *J. comp. Neurol.* **165**, 197-208 (1976).
- Eng, L. F., Kosek, J. C., Forno, L., Deck, J. & Bigbee, J. *Trans. Am. Soc. Neurochem.* **7**, 211 (1976).
- Kessler, D., Levine, L. & Fassman, G. *Biochemistry* **7**, 758-764 (1968).
- Zuckerman, J., Herschman, H. R. & Levine, L. *J. Neurochem.* **17**, 251-256 (1970).
- Herschman, H. R., Levine, L. & deVellis, J. *J. Neurochem.* **18**, 629-633 (1971).
- Pfeiffer, S. E., Herschman, H. R., Lightbody, J. & Sato, G. *J. cell. Physiol.* **75**, 329-340 (1970).
- Benda, P., Lightbody, J., Sato, G., Levine, L. & Sweet, W. *Science* **161**, 370-371 (1968).
- Moore, B. *Proteins of the Nervous System* (Raven, New York, 1973).
- Irie, R., Irie, K. & Morton, D. *Cancer Res.* **36**, 3510-3517 (1976).
- Irie, R., Giuliano, A. E. & Morton, D. *J. natn. Cancer Inst.* **63**, 367-373 (1979).
- Sarich, U. M. & Wilson, A. C. *Science* **154**, 1563-1564 (1966).
- Nonno, L., Herschman, H. R. & Levine, L. *Archs Biochem. Biophys.* **136**, 361-367 (1969).
- Fabricant, R., DeLarco, J. & Todaro, G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 565-569 (1977).
- Levine, L. in *Handbook of Experimental Immunology* (ed. Weir, D. M.) (Blackwell, Oxford, 1967).
- Zuckerman, J., Herschman, H. R. & Levine, L. *J. Neurochem.* **17**, 251-256 (1970).

Polar redistribution of ¹²⁵I-labelled insulin on the plasma membrane of cultured human lymphocytes

Philippe Barazzone*, Jean-Louis Carpentier*, Phillip Gorden†, Emmanuel Van Obberghen† & Lelio Orci*

* Institute of Histology and Embryology, School of Medicine, University of Geneva, Geneva, Switzerland

† Diabetes Branch, National Institutes of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014

Cultured human lymphocytes of the IM-9 line have specific binding sites for ¹²⁵I-insulin that have been characterized in detail^{1,2}. When the hormone-cell interaction was studied by quantitative electron microscopic (EM) autoradiography, it was shown that initial binding was to the plasma membrane and that at 37 °C a small fraction of the ligand was internalized by the cell^{3,4}. The internalized ligand was found preferentially localized in lysosomes in the Golgi region of the cell⁵. It is well known that di- or multivalent ligands redistribute in the plane of the membrane before internalization⁶. We now report that following initial binding the univalent ligand insulin also undergoes a polar redistribution ('capping') before internalization.

Cultured human lymphocytes have a polarity that can be distinguished in appropriately taken photomicrographs of thin sections of these cells (Fig. 1). The Golgi pole was arbitrarily assigned to half the cell and the opposite pole was assigned to the other half as indicated in Fig. 1.

In this study we have assumed that initial receptor binding is most accurately reflected at low temperature and at early times of incubation. When 0.2 nM ¹²⁵I-insulin was incubated at 15 °C for 2 min and the hormone-cell interaction examined by EM autoradiography, grains became distributed on the plasma membrane of the cell in a statistically indistinguishable fashion (Table 1). We interpret this to mean that insulin receptors or groups of receptors are initially distributed at each pole of the cell with equal frequency.

Table 1 Distribution of autoradiographic grains on the surface of IM-9 lymphocytes during incubation with 0.2 nM ¹²⁵I-insulin

	2 min, 15 °C	90 min, 15 °C	2 min, 37 °C	60 min, 37 °C
Golgi pole	53%	58%	72%	73%
Opposite pole	47%	42%	28%*	27%*

Number of grains analysed = 2 min, 15 °C: 148; 90 min, 15 °C: 154; 2 min, 37 °C: 140; 60 min, 37 °C: 164.

* As determined by the χ^2 test these grain distributions are significantly different ($P < 0.005$) from the grain distribution found after 2 min of incubation at 15 °C.

We next investigated whether this was true for insulin at physiological concentrations. Labelled insulin was, therefore, incubated with cultured lymphocytes for a longer period of time at 15 °C and 37 °C. By 90 min at 15 °C there was no statistically significant difference between the percentage of grains associated with each pole of the cell (Table 1). By contrast, by 2 min of incubation at 37 °C there was a preferential association of the ligand with the Golgi pole of the cell and the same distribution of the labelled material was apparent after 60 min of incubation (Table 1, Fig. 2). We have found that although the affinity of insulin for its receptor is decreased with increasing temperature, there is no change in the total number of receptors when the incubation temperature is raised from 15 to 37 °C (unpublished observations). We interpret the present data to mean that, although insulin initially binds with equal frequency to either

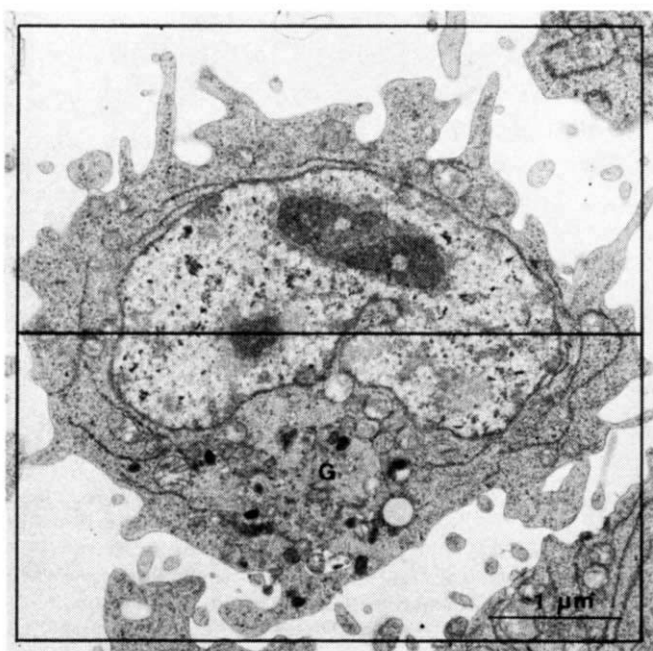


Fig. 1 The method used to determine the two plasma membrane segments of the lymphocytes for evaluation of ^{125}I -insulin redistribution (examples of lymphocytes with autoradiographic grains are shown in refs 3, 5, Figs 2, 6, respectively). Typically, the lymphocyte has an eccentric, indented nucleus. The Golgi area (G) lies close to the indentation of the nucleus. A square divided in two equal rectangles by a horizontal line was superimposed on each cell so that the Golgi area fitted approximately the centre of one of the rectangles. The plasma membrane segment contained within that rectangle was considered as the 'Golgi pole' of the membrane; the number of autoradiographic grains present within ± 250 nm of the plasma membrane in each rectangle was then recorded separately. Approximately 100 cells showing the typical morphology illustrated were evaluated in each experimental condition. Lymphocytes were incubated with 0.2 nM ^{125}I -insulin at 15 and 37°C as described elsewhere^{3,5}. At each time point selected cells were fixed with 4% phosphate-buffered (pH 7.4) glutaraldehyde, dehydrated, embedded in Epon and thinly sectioned. High-resolution autoradiography of the thin sections was carried out using Ilford L4 emulsion and Microdol-X developer after 3–5 weeks' exposure time^{3,5}.

pole of the cell, at 37°C there is a rapid redistribution of the ligand in the plane of the membrane.

It has previously been shown in human fibroblasts that some ligands initially bind preferentially to specialized regions of the membrane known as coated pits^{7–11}. The lymphocyte differs from the fibroblast in that the coated regions of the plasma membrane are much less marked with respect to the size of the coated segment and coat development. Although this factor must be taken into account in the analysis of our data, we find no initial preferential localization of labelled insulin to coated segments of the membrane after 2 min of incubation at either 15 or 37°C . These data are similar to our findings of insulin binding in isolated hepatocytes, to those of Billheimer *et al.* for low-density lipoprotein (LDL) binding in circulating monocytes and to those of Haigler *et al.* studying epidermal growth factor (EGF) binding in A-431 epithelioid carcinoma cells^{12–14}, but

may not rule out a participation of coated segments of the membrane in internalization of the labelled material.

Schlessinger *et al.*¹⁵ have shown in cultured mouse fibroblasts that, at high concentration, insulin conjugated to lactalbumin-fluorescein undergoes patching before internalization. Whether a similar process occurs in our cultured lymphocytes cannot be determined from these studies. The redistribution of insulin in the plane of the membrane to the Golgi pole of the cultured human lymphocytes seems more closely analogous to the phenomenon of capping. Taken with the data of Schlessinger *et al.*¹⁵, the present results suggest that insulin binding to the cell surface induces a patch-cap formation similar to that described for di- or multivalent ligands⁶. In this respect, it is interesting that the capping of ^{125}I -insulin on the cell surface is qualitatively and, most importantly, quantitatively identical to that observed for a unique immunoglobulin (anti-insulin receptor antibody) which binds specifically to the insulin receptor¹⁶, and which was isolated from a patient with a severe insulin resistance.

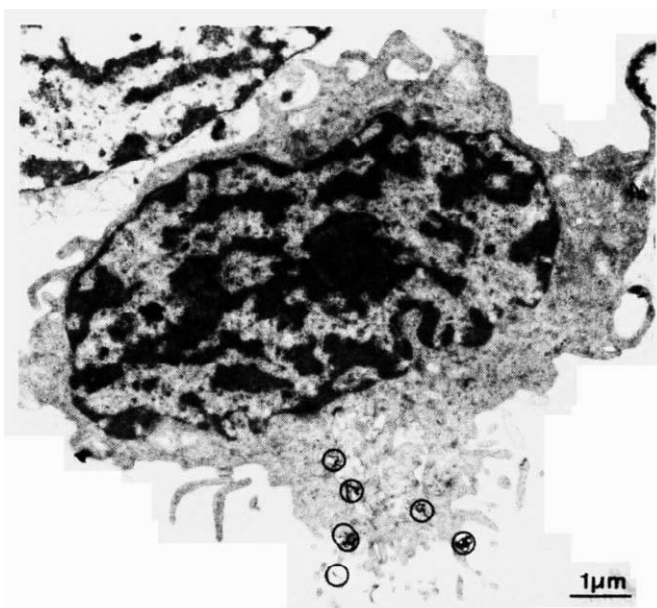


Fig. 2 IM-9 lymphocytes incubated for 60 min at 37°C with 0.2 nM ^{125}I -insulin. Autoradiographic grains (encircled) are concentrated in the region of the cell characterized by numerous microvilli and membrane infoldings (uropod). This region corresponds to the Golgi pole of the lymphocyte.

If one assumes that at physiological concentrations insulin behaves as a univalent ligand, it is then clear that bivalency of the ligand is unnecessary either for mobility in the plane of the membrane or for internalization. However, whether the ligand induces some form of aggregation or cross-linking of its receptor remains unknown.

Finally, we have proposed that endocytosis may be the major mechanism by which polypeptide hormone receptors are lost from the cell surface⁵. Micro-aggregation, lateral mobility and internalization of the hormone and its specific receptor in the plane of the membrane could provide a mechanism for selective loss of a single class of receptors analogous to the removal of specific immunoglobulins from the surface of lymphocytes¹⁷.

This work was supported by grant 3.120.77 from the Swiss NSF. During this work P. Gorden was a visiting professor at the Institute of Histology and Embryology in Geneva.

Received 4 March; accepted 14 May 1980.

- Gavin, J. R. III, Gorden, P., Roth, J., Archer, J. A. & Buell, D. N. *J. biol. Chem.* **248**, 2202 (1973).
- De Meyts, P. in *Methods in Receptor Research* (ed. Blecher, M.) 301 (Dekker, New York, 1976).
- Carpentier, J.-L. *et al. J. clin. Invest.* **61**, 1057 (1978).
- Goldfine, I. D., Jones, A. L., Hradek, G. T., Wong, K. Y. & Mooney, J. S. *Science* **202**, 760 (1978).
- Gorden, P. *et al. J. Cell Sci.* **39**, 77 (1979).

Table 2 Relationship of autoradiographic grains to coated regions of the plasma membrane

	2 min, 15°C	2 min, 37°C
Total no. of grains examined	252	292
No. of grains associated with coated invaginations	5 (2%)	6 (2.1%)

The criteria for selection were: (1) the grain centre was ± 250 nm from the plasma membrane, and (2) the plane of the section was appropriate for a clear identification of the plasma membrane. A grain was considered associated with a coated invagination if its centre was closer than 250 nm from a coated invagination. These methods are identical to those previously used for analysis of binding of ^{125}I -EGF and ^{125}I -LDL to cultured fibroblasts^{9,10}.

6. de Petris, S. in *Dynamic Aspects of Cell Surface Organization* Vol. 3 (eds Poste, G. & Nicholson, G. L.) 643 (North-Holland, Amsterdam, 1977).
7. Anderson, R. G. W., Brown, M. S. & Goldstein, J. L. *Cell* **10**, 351 (1977).
8. Orci, L. *et al. Expl Cell Res.* **113**, 1 (1978).
9. Carpentier, J.-L. *et al. Expl Cell Res.* **121**, 135 (1979).
10. Gorden, P., Carpentier, J.-L., Cohen, S. & Orci, L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5025 (1978).
11. Willingham, M. C., Maxfield, F. R. & Pastan, I. M. *J. Cell Biol.* **82**, 614 (1979).
12. Carpentier, J.-L., Gorden, P., Freychet, P., Le Cam, A. & Orci, L. *J. clin. Invest.* **63**, 1249 (1979).
13. Billheimer, D. N., Ho, Y. V., Brown, M. S., Anderson, R. G. W. & Goldstein, J. L. *J. clin. Invest.* **61**, 678 (1978).
14. Haigler, M. T., McKanna, J. A. & Cohen, S. *J. Cell Biol.* **81**, 382 (1979).
15. Schlessinger, J., Schechter, T., Willingham, M. C. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2659 (1978).
16. Carpentier, J.-L., Van Obberghen, E., Gorden, P. & Orci, L., in preparation.
17. Taylor, R. B., Duffus, M. R. M., Raff, M. C. & De Petris, S. *Nature new biol.* **233**, 225 (1971).

Increased human red cell cation passive permeability below 12 °C

G. W. Stewart, J. C. Ellory & R. A. Klein*

Physiological Laboratory, Downing Street, Cambridge CB2 3EG, UK

* Medical Research Council, Molteno Institute, University of Cambridge, Cambridge CB2 3EE, UK

The rate of most biological reactions declines as the temperature is reduced; indeed, cooling is often used to limit or terminate reactions. We report here a paradoxical temperature response of red cell K and Na permeability below 12 °C. This may be interpreted thermodynamically in terms of a membrane-ordering phenomenon, an observation supported by a variety of other physical measurements on red cell membranes reported in the literature.

Surprisingly few data are available on the effects of temperature on Na and K transport in human erythrocytes. This may be due in part to the presence of at least three components of Na and K movement in human red cells. These can be resolved as the ouabain-sensitive Na-K pump^{1,2}, the furosemide-sensitive Na-K co-transport system³ and a residual process showing linear concentration dependence¹. For normal human red cells, at 5 mM external K, K influx is roughly divided between these three processes in the proportion 60%, 35% and 5%, respectively, at 37 °C. We have measured the temperature dependence of the three components of K uptake over the temperature range 4–37 °C (Fig. 1). The two K transport systems which show saturable kinetics and which can be identified using known inhibitors (ouabain for the Na-K pump and furosemide, or other loop diuretics such as bumetanide, for the Na-K co-transport system), fell off sharply with decreasing temperature, no significant fluxes being measurable below about 4 °C (Fig. 1). In marked contrast, however, the residual K influx, which shows a

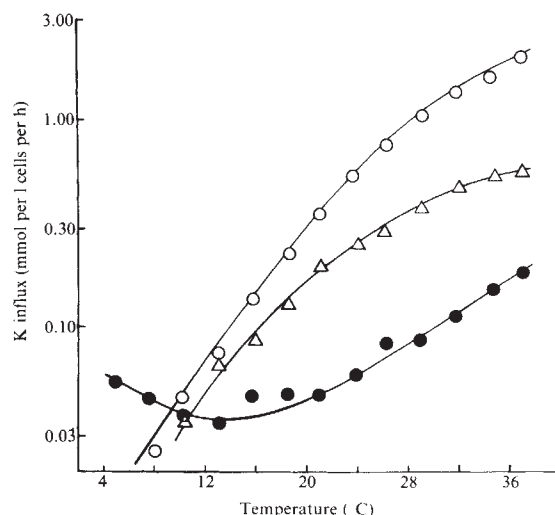


Fig. 1 The three components of K influx as a function of temperature. K influx was measured by previously described methods¹⁵ using ⁴²K or tracer ⁸⁶Rb (ref. 16) in a 142 mM NaCl, 7.5 mM KCl, 15 mM MOPS (Sigma) and 5 mM glucose solution, pH 7.4, at 20 °C, 7.3 at 0 °C, 7.48 at 37 °C. Inhibitors (ouabain and furosemide, Hoechst) were added to final concentrations of 0.1 mM and 1 mM, respectively. Cells were preincubated at 37 °C for 15 min before cooling to allow binding. The cells were exposed to isotope for up to 3 h before rapid washing four times in ice-cold MgCl₂ 106 mM and Tris 10 mM. ○, Ouabain-sensitive; △, furosemide-sensitive; ●, ouabain/furosemide-insensitive.

linear K concentration dependence and is ouabain- and furosemide-resistant, reached a minimum at about 12 °C, below which temperature the flux began to increase again. In three experiments in which this flux was measured at 0 °C and 10 °C, the ratio K influx 0 °C: 10 °C was 2.1 ± 0.5 (s.e.m.). Ouabain- and furosemide-resistant Na efflux (Fig. 2), K efflux and Na influx (data not shown) all showed a similar paradoxical pattern of temperature dependence with a minimum at about 12 °C. In a second series of experiments, net Na accumulation was measured following long-term (22 h) incubation in media containing ouabain and bumetanide (used as an alternative to furosemide because of its greater potency). The results (Table 1) demonstrate that Na accumulation was significantly lower following incubation at 10 °C than with either 20 °C or 0 °C. This experiment, therefore, confirms the observations from tracer fluxes that there was a minimum around 10 °C for Na permeability, as also shown by Dalmak & Wieth¹⁷.

Although ouabain binding shows a very low reversibility at 0 °C (ref. 4), it was important to rule out the possibility of reversal of bumetanide (or furosemide) binding at low temperatures. Cells were always pretreated with maximal concentrations of inhibitors for 15 min at 37 °C before flux measurements, and as the K concentration dependence of K influx, measured at 0, 4, 11 and 37 °C, was found always to be linear, reversal of inhibition at low temperatures, which would restore a saturable component of K influx, can be discounted. Possible participation of the anion transport protein in this effect (through, for example NaCO₃⁻ or KCO₃⁻ transport)⁵ is unlikely because the anion exchange flux rate is markedly inhibited by temperature between 10 and 0 °C (ref. 6). In any case 0.1 mM SITS (4-acetamido-4'-isothiocyanatostilbene-2,2'-disulphonic acid, sodium salt, BDH), a potent inhibitor of the anion transport system⁷, had no effect. Membrane potential effects can be ruled out because both influx and efflux tracer cations were increased at low temperatures, and there was no significant temperature-dependent change in membrane potential (measured by the dis-C₃-(5) dye method⁸) in ouabain- and furosemide-treated cells.

In the present work, we have shown that the linear components of Na and K fluxes have minimum values at around

Table 1 Changes in internal erythrocyte [Na] after storage as a function of temperature

	[Na] _i μmol per g Hb	Difference
Before storage	31.9 ± 0.1	
After storage		
20 °C	51.8 ± 0.3	19.9 ± 0.4
10 °C	46.7 ± 0.2	14.8 ± 0.2*
0 °C	52.0 ± 0.3	20.1 ± 0.4*

After washing 4× in (mM) NaCl 145, KCl 5, MOPS 15 and glucose 5, pH 7.4, at 20 °C, followed by 15-min incubation with inhibitors at 37 °C, cells were incubated for 22 h at 4% haematocrit at the temperatures shown with ouabain and bumetanide 10⁻⁴ M each, with continuous gentle mixing. After storage, the cells were washed 4× in 106 mM MgCl₂ and 10 mM Tris, packed and lysed in water. [Na] was measured by flame photometry and haemoglobin (Hb) was measured with Drabkin's reagent at 540 nm. The results are expressed per g Hb (± s.e.m., n = 6) to exclude any effects of temperature on cell volume.

*Significantly different at $P < 0.001$, Student's *t*-test.

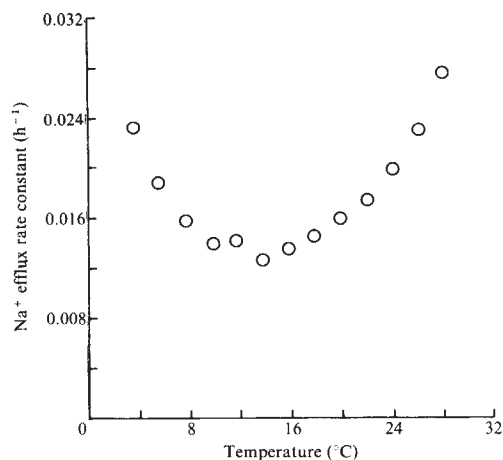


Fig. 2 Na efflux as a function of temperature. Fresh human cells were loaded with ^{24}Na as previously described¹⁵. Ouabain (10^{-4} M) and bumetanide (10^{-4} M) were added towards the end of the loading incubation to allow binding before rapid washing in cold 145 mM NaCl, 5 mM KCl, 15 mM MOPS and 5 mM glucose solution, containing the inhibitors.

12 °C. Wieth⁹ has reported a temperature-dependent stimulation of the ouabain-insensitive Na and K fluxes in human erythrocytes by the anions SCN^- and salicylate (used at 100 mM concentration) which gave rise to a flux minimum at about 18 °C. In thermodynamic terms such paradoxical behaviour could be the result of two mechanisms: the summation of two processes with enthalpies of activation of opposite sign (for example, Wieth⁹ suggested that the ions SCN^- and salicylate titrated fixed membrane charges exothermically at low temperatures) or the result of a single process whose entropy change is a function of temperature. According to Eyring's theory¹⁰, the temperature dependence of diffusion (D) is given by

$$D = cTe^{-\Delta G^*/RT}$$

where c is a constant, T the absolute temperature, R the gas constant and ΔG^* the Gibbs free energy change for activation of the transition state. From the Gibbs equation $\Delta G^* = \Delta H^* - T\Delta S^*$ where ΔH^* and ΔS^* are, respectively, the enthalpy and entropy changes for activation. In the first case, ΔS^* is invariant with temperature. If $\ln D$ is differentiated with respect to $1/T$, one obtains

$$\frac{d \ln D}{d(1/T)} = -T - \frac{\Delta H^*}{R}$$

If $T \ll \Delta H^*/R$, this reduces to

$$\frac{d \ln D}{d(1/T)} = -\frac{\Delta H^*}{R}$$

the Arrhenius equation. If, in the second case, ΔS^* is a function of temperature denoted by $\Delta S^*(1/T)$, the expression becomes

$$\frac{d \ln D}{d(1/T)} = \frac{d}{d(1/T)} \left\{ \frac{\Delta S^*(1/T)}{R} \right\} - \frac{\Delta H^*}{R} - T$$

The function $\ln D$ (and therefore D) can show a minimum if at some temperature

$$\frac{d}{d(1/T)} \Delta S^*(1/T) = \Delta H^* + RT$$

This equation expresses the thermodynamic condition necessary for a minimum to be observed in the relationship between ion flux and temperature, consistent with the present data. A temperature-dependent change in the entropy of activation for the transition complex could result from a change in membrane ordering. In this context workers using physical techniques to

study membrane ordering have reported relevant observations: (1) Raman spectroscopy (Verma and Wallach¹¹) showed discontinuities in lipid CH-stretching and membrane-bound β -carotene intensities in the temperature range -4 °C to 20 °C; (2) Tanaka and Ohnishi¹² demonstrated an inflection point at 18 °C in the phosphatidylcholine spin label line splitting for erythrocyte ghosts; (3) in ether-extracted erythrocyte ghosts, Cullis¹³ found a change in ^{32}P nuclear magnetic resonance linewidth at about 20 °C; (4) using a fluorescence probe to assess membrane fluidity, Kapitza and Sackmann¹⁴ observed a marked increase in this parameter between 12 and 17 °C. All these methods reflect membrane ordering and therefore entropy. However, as the temperature of the flux minimum is a function not only of $d/d(1/T)\Delta S^*(1/T)$ but also of ΔH^* and RT , points of inflection identified from these data cannot be simply related to the flux minimum described here.

Our flux measurements may therefore represent a functional correlate of the temperature-dependent changes in membrane ordering evidenced by other physical methods.

We thank Dr T. J. Rink for measuring membrane potential, Dr V. L. Lew for helpful discussions, and the Wellcome Trust for supporting G.W.S.

Received 13 February; accepted 16 May 1980.

- Glynn, I. M. *J. Physiol., Lond.* **134**, 278-310 (1956).
- Post, R. L. & Jolly, P. C. *Biochim. biophys. Acta* **25**, 118-128 (1957).
- Wiley, J. S. & Cooper, R. A. *J. clin. Invest.* **53**, 745-755 (1974).
- Joiner, C. H. & Lauf, P. K. *J. Membrane Biol.* **21**, 99-112 (1975).
- Funder, J. & Wieth, J. O. *Alfred Benzon Symp.* **14**, 520-530 (eds Lassen, U., Ussing, H. H. & Wieth, J. O.) (Munksgaard, Copenhagen, 1980).
- Dalmark, M. & Wieth, J. O. *J. Physiol., Lond.* **224**, 583-610 (1972).
- Knauf, P. & Rothstein, A. *J. gen. Physiol.* **58**, 190-210 (1971).
- Hladky, S. B. & Rink, T. J. *J. Physiol., Lond.* **263**, 287-319 (1976).
- Wieth, J. O. *J. Physiol., Lond.* **207**, 563-580 (1970).
- Stein, W. D. *The Movement of Molecules across Cell Membranes* (Academic, London, 1967).
- Verma, S. P. & Wallach, D. F. H. *Biochim. biophys. Acta* **436**, 307-318 (1975).
- Tanaka, K. & Ohnishi, S. *Biochim. biophys. Acta* **426**, 218-231 (1975).
- Cullis, P. R. *FEBS Lett.* **68**, 173-176 (1976).
- Kapitza, H. G. & Sackmann, E. *Biochim. biophys. Acta* **595**, 56-64 (1980).
- Ellory, J. C. & Dunham, P. B. *Alfred Benzon Symp.* **14**, 409-423 (eds Lassen, U., Ussing, H. H. & Wieth, J. O.) (Munksgaard, Copenhagen, 1980).
- Fortes, P. A. G., Ellory, J. C. & Lew, V. L. *Biochim. biophys. Acta* **318**, 262-272 (1973).
- Dalmark, M. & Wieth, J. O. *Biochim. biophys. Acta* **219**, 525-527 (1970).

Plateau potentials in pancreatic islet cells are voltage-dependent action potentials

Daniel L. Cook, Wayne E. Crill & Daniel Porte Jr

Veterans Administration Hospital, 4435 Beacon Avenue So., Seattle, Washington 98108

Islet cells stimulated by a steady glucose concentration above a certain threshold (usually 5-7 mM) display cyclic membrane potential changes¹. This islet electrical activity consists of action potentials (spikes) superimposed on a depolarized plateau (see Fig. 1). Higher levels of glucose prolong the plateau phase and shorten the silent phase until the plateau potentials merge to produce continuous spiking at 20-25 mM. The fraction of each cycle spent in the plateau phase correlates with glucose concentration and insulin release rate², suggesting that control of the onset and offset of the plateau is important in the regulation of insulin release. The abrupt transitions between silent and plateau phases suggest that the plateau is a persistent, voltage-dependent action potential as in cardiac muscle. This idea is supported by evidence³ that Ca^{2+} influx is stimulated by K^+ depolarization of islets. However, this voltage dependency can be explained by electrical evidence^{4,5} that the spikes are generated by voltage-dependent Ca^{2+} currents. Furthermore, there is evidence against the voltage dependency of the plateau because it seemed to be unaffected by intracellular current injection and K^+ depolarization, which clearly affected spik-

ing^{5,6}. As there are alternative explanations (see below) for these last negative findings, we have re-examined this important question using a new technique of direct electrical stimulation of entire, isolated islets using a suction electrode. Contrary to the previous negative findings, our present results clearly show that the plateau potential is a voltage-dependent, persistent action potential.

Pieces of pancreas from Swiss-Webster mice fed *ad libitum* were perfused in a Krebs-Ringer bicarbonate buffer gassed with O₂/CO₂ (95:5) and maintained at 37 °C. Individual islets (100–300 µm diameter) were microdissected free of exocrine tissue and held by the hollow glass electrode (90–200 µm inside tip diameter) with gentle suction (5–10 mm H₂O). A stimulus isolation unit provided constant current stimuli through this suction electrode. Standard electrophysiological techniques with glass microelectrodes (tip resistance 300–400 MΩ) were used to record islet cell membrane potentials. Because the suction electrode current passes through the entire islet, extracellular potentials within the islet cannot be assumed equal to the bath ground potential. Therefore, potentials measured during stimulation are not a true measure of transmembrane potential. Glucose-responsive islet cells were identified by their typical pattern of electrical activity and were used only if recordings were stable for at least 10 min. These records typically had silent phase potentials between –50 and –60 mV, and at the time of recording the islets had been exposed to 11.1 or 13.7 mM glucose for at least 30 min.

Figure 1, line 1, shows the rhythmic electrical response to 11.1 mM glucose. In line 2, this regular endogenous rhythm was interrupted by an 8-µA, 2-s-long depolarizing current (from the suction electrode) which evoked an all-or-none plateau potential indistinguishable from endogenous plateaus. The ensuing electrical rhythm is identical to that observed before the stimulus but is shifted as if a pacemaker had been reset. Line 3 of Fig. 1 demonstrates that this all-or-none response has a definite electrical threshold, as a slightly weaker stimulus (6 µA) did not initiate a plateau and had no effect on the plateau rhythm.

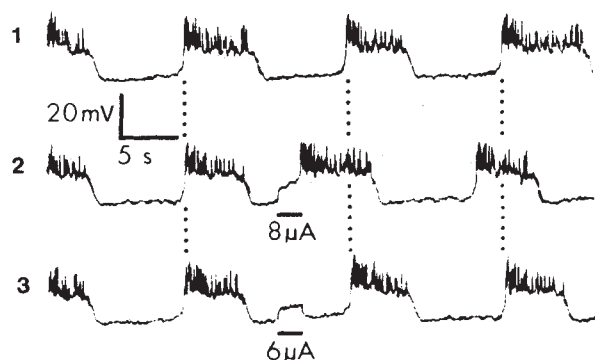


Fig. 1 Responses to depolarizing currents. Three excerpts from a membrane potential recording from a single islet cell in the presence of glucose (11.1 mM). Line 1, control activity; line 2, a premature plateau potential initiated by an 8-µA, 2-s-long depolarizing current presented 2 s into the silent phase; line 3, failure of a similar 6-µA stimulus to initiate a plateau.

Plateau termination is also voltage dependent because an endogenous plateau can be aborted by a brief (100 ms) hyperpolarizing stimulus current of –18 µA (Fig. 2, line 2). This all-or-none response has the same slow time course of repolarization and length of silent phase as seen in the endogenous activity. There is also a sharp stimulus threshold, as a slightly weaker current (–15 µA) does not interrupt the plateau (Fig. 2, line 3). Although not clearly shown in Fig. 2, artificial termination of plateaus interrupted and reset the endogenous rhythm as effectively as artificial plateau initiation (Fig. 1).

For both the initiation and termination of plateaus in all cells tested, we have consistently observed a reciprocal relationship

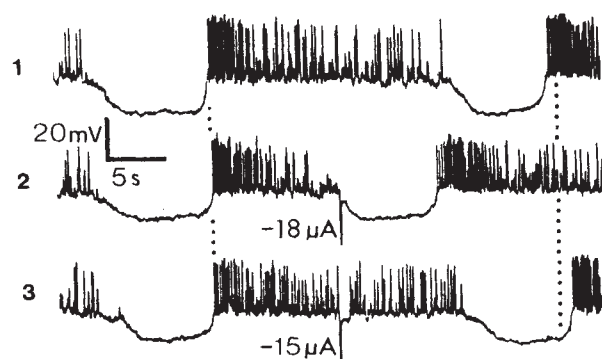


Fig. 2 Responses to hyperpolarizing currents. Membrane potential records from a second cell in the presence of glucose (13.7 mM). Line 1, control electrical activity; line 2, endogenous plateau prematurely terminated by a –18-µA, 100-ms-long hyperpolarizing current presented 10 s after the beginning of the plateau; line 3, failure of a similar –15-µA current to terminate a plateau.

between stimulus current strength and stimulus duration (data not shown), as is characteristic of excitability phenomena in other systems. For example, in one cell, a 10-µA current required 2 s to trigger a premature plateau whereas a 20-µA current required only 1 s. Similar relationships were found for termination of plateaus, although typically, plateau initiation has required longer stimuli than plateau termination (1–2 s compared with 50–100 ms).

Previous studies^{5,6} found that neither plateau length nor frequency was affected by stimulus currents passed through the recording electrode or by elevation of bath K⁺ concentration. These results probably reflect the difficulty of driving a coupled⁷ population of cells by stimulating a single cell, as well as accommodation to slow changes in K⁺. More rapid (1–2 s) bath K⁺ changes elicit premature plateaus (unpublished results) as readily as electrical stimulation through the suction electrode.

It is possible that our electrical stimuli affect plateau potentials indirectly by releasing endogenous islet neurotransmitters. However, this is unlikely because it seems to require that depolarizing stimuli selectively release acetylcholine (known to depolarize islet cells)⁸ whereas hyperpolarizing stimuli selectively release catecholamines (known to hyperpolarize islet cells)⁹. A more likely explanation is that the electrical stimuli are directly activating and deactivating a voltage-dependent membrane conductance. This conductance change underlies the prolonged plateau depolarization which drives the more rapid spiking activity. Thus, the islet cell plateau potential resembles voltage-dependent, persistent action potentials seen in myocardial cells¹⁰ and the slow potentials of molluscan pacemaker neurones¹¹. The underlying ionic mechanism in islet cells is unclear but may be due to increased conductance to Na⁺ or Ca²⁺ as in these other systems or to a decreased conductance to K⁺, as proposed for islet cells⁵.

The second key observation in the present work is that electrical stimuli interrupt and completely reset the rhythm of islet cell electrical activity. This seems to rule out electrically insensitive pacemaker mechanisms such as autonomous biochemical oscillators which merely drive membrane events. Furthermore, the pacemaker mechanism must depend on either the membrane potential directly or the fluxes of ions underlying the voltage changes. For instance, as in heart cells and molluscan neurones, one or more slow, voltage-dependent K⁺ conductances may serve as the pacemaker. Alternatively, islet cells are known to possess an electrogenic Na⁺–K⁺ pump¹² and a Ca²⁺-dependent, K⁺ conductance¹³. Thus, intracellular accumulation of Na⁺ or Ca²⁺ during the active plateau phase could, by either mechanism, activate a hyperpolarizing current which would

eventually trigger plateau repolarization. Depletion of the cation pool during the silent phase would lead to slow depolarization and the next plateau.

Although the specific ionic mechanisms of the plateau conductance change and the pacemaker mechanism remain unclear, these data conclusively show the islet cell plateau potential to be a persistent, voltage-dependent action potential whose pacing must involve membrane potential or ion fluxes.

We thank H. P. Meissner and I. Atwater for technical advice. This work was supported by the US Veterans Administration and NIH research grants AM 12829 and 21185. D.L.C. was supported by a Juvenile Diabetes Foundation post-doctoral fellowship.

Received 24 March; accepted 13 May 1980.

1. Dean, P. M. & Matthews, E. K. *Nature* **219**, 389–390 (1968).
2. Meissner, H. P. & Schmelz, H. *Pflügers Arch. Ges. Physiol.* **357**, 195–206 (1974).
3. Henquin, J. C. & Lambert, A. E. *J. Physiol., Paris* **72**, 747–756 (1976).
4. Matthews, E. K. & Sakamoto, Y. *J. Physiol., Lond.* **246**, 421–437 (1975).
5. Atwater, I., Ribalet, B. & Rojas, E. *J. Physiol., Lond.* **278**, 117–139 (1978).
6. Cook, D. L., Porte, D. Jr & Crill, W. E. (abstr) *10th Cong. int. Diabetes Fedn*, 46 (Int. Cong. Ser. 481, Excerpta Medica, Amsterdam, 1979).
7. Meissner, H. P. *Nature* **262**, 502–504 (1976).
8. Gagerman, E., Idahl, L.-A., Meissner, H. P. & Taljedal, I.-B. *Am. J. Physiol.* **235**, E493–E500 (1978).
9. Dean, P. M. & Matthews, E. K. *J. Physiol., Lond.* **210**, 255–264 (1970).
10. Reuter, H. A. *Rev. Physiol.* **41**, 413–424 (1979).
11. Gola, M. in *Abnormal Neuronal Discharges* (eds Chalazonitis, N. & Boisson, M.) 243–261 (Raven, New York, 1978).
12. Meissner, H. P. *J. Physiol., Paris* **72**, 757–767 (1976).
13. Atwater, I., Dawson, C. M., Ribalet, B. & Rojas, E. *J. Physiol., Lond.* **288**, 575–588 (1979).

Linkage relationship between variable and constant region allotypic determinants of human immunoglobulin heavy chains

J. P. Pandey*, E. Tung*, S. Mathur*, K. K. Nambodiri†, A. C. Wang*, H. H. Fudenberg*, W. A. Blattner‡, R. C. Elston† & C. G. Hames§

*Department of Basic and Clinical Immunology and Microbiology, Medical University of South Carolina, Charleston, South Carolina 29403; †Department of Biostatistics, The University of North Carolina, Chapel Hill, North Carolina 27514; ‡National Institutes of Health, Bethesda, Maryland 20014; §Evans County Health Department, 2 North Newton Street, Claxton, Georgia 30417

Allotypic (genetic) variations in the variable (V) region of immunoglobulin chains have been extensively studied in the domestic rabbit^{1–3}. The first allotypic determinant identified in the V region of human immunoglobulins, described in our laboratory, is designated Hv(1). Its detection, using protein McE and a rabbit antiserum to McE, and its mode of inheritance based on family and population studies, have been reported previously⁴. Briefly, Hv(1) is located in the V region of human immunoglobulin heavy (H) chains of G, M and A classes. It is inherited as an autosomal dominant trait, with a gene frequency in white people of 0.189 ± 0.023 . The polymorphic nature of Hv(1) has also been confirmed in black people, studies on a large black kindred ($n=80$) having confirmed its autosomal dominant mode of inheritance. The gene frequency (determined by the Elston–Stewart method using maximum likelihood⁵) in blacks is 0.278 ± 0.067 . Hv(1) seems analogous to murine idiotypes of the so-called *I*_{dx} category⁶ and to rabbit V_H-region allotypes in that it is not restricted to any particular immunoglobulin class. This observation is consistent with our previous finding that the same V-region amino acid sequence is able to associate with constant (C) regions of different H-chain classes⁷. However, an important difference between Hv(1) and the idiotype markers in mice and rabbits is that the former is presumably associated with the framework residues whereas the latter are associated primarily with the hypervariable regions. We now report results obtained from family studies which indicate absence of linkage between Hv(1) and the human C-region markers Gm (human IgG H chains) and Km (κ -type light chains). These results provide new information about the genetic control of immunoglobulin synthesis in man.

Serum samples from one large black kindred ($n=80$) and one large white kindred ($n=220$) and from three smaller white families ($n=84$) were typed for Hv(1), Km(1) and nine Gm markers (1, 2, 3, 5, 6, 13, 14, 17, 21) by our standard haemagglutination inhibition method^{4,8}. All sera were diluted 1:16 for detection of Gm and Km markers. The Hv(1) marker

was detected by the use of serial dilutions, with inhibiting titres of 1:16 or greater considered as positive. The lod score method of Morton⁹ was used to test for linkage between Hv(1) and the C-region markers, basing the likelihood on the complete pedigree structures^{5,10}. This method takes advantage of available data from every sort of informative test cross, that is, from double intercrossovers and single backcrosses as well as double backcrosses. It determines the relative odds that a sequence of genotypes among siblings resulted from linked loci with a certain percentage of meiotic crossing over, as compared with the possibility that the sequence occurred by chance from unlinked loci. For convenience, the ratio of these two probabilities—recombination fraction $\theta < 1/2$ (linkage) compared with $\theta = 1/2$ (no linkage)—is expressed as its logarithm. The $\log_{10}$ of this relative probability is called the log of the odds or 'lod score', and the value of θ which corresponds to the maximum value of the lod score is taken as the estimate of the recombination fraction θ . Computer program LIPED¹⁰ was used to calculate lod scores. The lod scores for linkage between Hv(1) and Gm and Km markers were negative (Table 1). (A negative value argues against linkage, a lod score greater than 2 is suggestive of linkage, and a value greater than 3 is accepted as proof of linkage.) On the basis of the data presented here, close linkage between Hv(1) and Gm and Km is excluded.

Investigations in humans, as well as mice and rabbits, suggest that unlinked genes code for H chains, κ -type light (L) chains and λ -type L chains^{11–14}. Absence of linkage between Hv(1) and Km is compatible with these observations. However, the absence of linkage between Hv(1) and Gm does not concur with findings in the rabbit and the mouse.

Most models of immunoglobulin genes predict close linkage between V- and C-region genes. V_H-region markers have been analysed in considerable detail with regard to linkage to C_H loci. Rabbit a-group allotypes, located on the V_H region, are linked to the d- and e-group allotypes located on the γ -chain of C_H (refs 15, 16), and in mice some idiotype (V-region) markers are linked to C_H allotypes^{6,17,18}. However, the results reported here are not entirely unprecedented. For example, one study of b-group allotypes and the LId idio type in rabbits gave no evidence of linkage between V_L and C_L genes¹⁹.

The simplest interpretation of the present finding is that Hv(1) is not a V-region allotype. However, extensive inhibition

Table 1 Lod scores for linkage between Hv(1) and Gm and Km markers

Recombination fraction (θ)*	Gm	Km
0.05	−9.93	−3.47
0.1	−6.33	−2.18
0.2	−2.72	−0.86
0.3	−0.99	−0.29
0.4	−0.19	−0.04

*Assuming that the recombination fractions (θ) in males and females are equal.

experiments involving Fab and Fc fragments and isolated H and L chains, reported previously⁴, demonstrated that the antigenic determinant for Hv(1) is located on the V region of human immunoglobulin heavy chains. Another interpretation of our results is that the Hv(1) and Gm phenotypes are controlled by two types of genes, structural genes coding for the antigenic determinant and regulatory genes controlling the expression of the determinant. Regulatory genes need not be linked to each other, nor to the genes they regulate²⁰. A number of observations in several species suggest that the classical mendelian view, in which allotypes are encoded by simple allelic structural genes, may be an oversimplification. For instance, in humans, discrepancy between the genotype and the expressed phenotype, and in rabbits, allotype suppression, latent allotypes and the 'pecking order' in the expression of allotypes, all suggest the existence of regulatory genes which control the expression of structural genes encoding immunoglobulin allotypes²¹⁻²⁷. Indeed, it has been suggested that expression of other genetic polymorphisms (for example, HLA and ABO) may also be under the control of regulatory genes²⁸.

Alternatively, the loci for Hv(1) and Gm may be on the same chromosome, but very far apart. In murine H and L chains, the V-region structural genes are located at some distance from the C-region structural genes in the germ-line DNA²⁹⁻³³. When two loci are very far apart, their genetic behaviour in family studies is often indistinguishable from that of independently segregating loci (that is, those on different chromosomes). One approach to this problem is genetic mapping by the somatic cell hybridization technique³⁴. Such studies are in progress in our laboratory.

We thank Vennie Moore for technical assistance, Charles Smith for editorial assistance, and Mary Marazita, John Thompson and Drs S. Hsia and T. Fogle for collection of some of the samples used in this study. This research was supported in part by USPHS grants HD-09938, CA-25746, RR-05767, MH-26721, HL-03341 and GM-16697 and NSF grant PCM-24809.

Received 20 December 1979; accepted 29 April 1980.

- Dray, S., Young, G. O. & Gerald, L. J. *Immun.* **91**, 403-415 (1963).
- Todd, C. W., *Biochem. biophys. Res. Commun.* **11**, 170-175 (1963).
- Nisonoff, A., Hopper, J. E. & Spring, S. B. *The Antibody Molecule*, 346-370 (Academic, New York, 1975).
- Wang, A. C. et al. *Science* **200**, 327-329 (1978).
- Elston, R. C. & Stewart, J. *Hum. Hered.* **21**, 523-542 (1971).
- Weigert, M. & Potter, M. *Immunogenetics* **4**, 401-435 (1977).
- Wang, A. C., Wang, I. Y. & Fudenberg, H. H. *J. biol. Chem.* **252**, 7192-7199 (1977).
- Vyas, G. N., Fudenberg, H. H., Pretty, H. M. & Gold, E. R. *J. Immun.* **103**, 274-279 (1968).
- Morton, N. E. *Am. J. hum. Genet.* **7**, 277-318 (1955).
- Ott, J. *Am. J. hum. Genet.* **26**, 588-597 (1974).
- Steinberg, A. G. & Matsumoto, H. *Hum. Biol.* **36**, 77-85 (1964).
- Mage, R., Lieberman, R., Potter, M. & Terry, W. D. in *The Antigens V.1* (ed. Sela, M.) 299-376 (Academic, New York, 1973).
- Natvig, J. B. & Kunkel, H. G. *Adv. Immun.* **16**, 1-60 (1973).
- Kindt, T. J. *Adv. Immun.* **21**, 35-86 (1975).
- Hamers-Casterman, C. & Hamers, R. *Immunogenetics* **2**, 597-603 (1975).
- Mage, R. G. *Annls Immun. Inst. Pasteur, Paris* **130C**, 105-114 (1979).
- Eichmann, K., Tung, A. S. & Nisonoff, A. *Nature* **250**, 509-511 (1974).
- Mäkelä, O. & Weigert, M. in *Immune System: Genetics and Regulation* (eds Sercarz, E. E., Herzenberg, L. A. & Fox, C. F.) 123-125 (ICN-UCLA Symp., Academic, New York, 1977).
- Sogn, J. A., Yarmush, M. L. & Kindt, T. J. in *Immune System: Genetics and Regulation* (eds Sercarz, E. E., Herzenberg, L. A. & Fox, C. F.) 91-97 (ICN-UCLA Symp., Academic, New York, 1977).
- McClintock, B. *Proc. natn. Acad. Sci. U.S.A.* **36**, 344-355 (1950).
- Lobb, N., Curtin, C. C. & Kidson, C. *Nature* **214**, 783-785 (1967).
- Rivat, L., Gilbert, D. & Ropartz, C. *Immunology* **24**, 1041-1049 (1973).
- Gutman, G. A., Loh, E. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5046-5050 (1975).
- Wang, A. C. *Fedn Eur. biochem. Soc. Symp.* **36**, 19-31 (1975).
- Strosberg, A. D. *Immunogenetics* **4**, 499-513 (1977).
- Strosberg, A. D. in *Immune System: Genetics and Regulation* (eds Sercarz, E. E., Herzenberg, L. A. & Fox, C. F.) 115-122 (ICN-UCLA Symp., Academic, New York, 1977).
- Yarmush, M. L., Sogn, J. A. & Kindt, T. J. *Annls Immun. Inst. Pasteur, Paris* **130C**, 143-156 (1979).
- Bodmer, W. F. *Transplant Proc.* **5**, 1471-1475 (1973).
- Brack, C., Hirowa, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1-14 (1978).
- Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370-375 (1979).
- Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 857-861 (1979).
- Davis, M., Early, P., Calame, K., Livant, D. & Hood, L. in *Eukaryotic Gene Regulation* (eds Axel, R., Maniatis, T. & Fox, C. F.) 393-406 (ICN-UCLA Symp., Academic, New York, 1979).
- Davis, M. M. et al. *Nature* **283**, 733-739 (1980).
- Croce, C. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3416-3419 (1979).

A test for the selective origin of environmentally correlated allozyme patterns

J. M. Malpica & J. M. Vassallo

Departamento de Genética, INIA, Carretera de La Coruña Km.7, Madrid-35, Spain

Differences in the kinetic and/or endurance properties of enzymes encoded by alternative alleles have been suggested as the basic level at which selection could act on allozyme polymorphisms^{1,2}. The correlation between spatial differences in gene frequencies and an environmental variable (hereafter designated cline, in its widest sense), when not spurious, has been proposed to show the working of selection³. Let us suppose that a given cline has originated through an environmentally imposed selection on enzyme activity. Because differences in enzymatic activities among individuals can be split into structural-gene and 'genetic-background' components⁴⁻⁶, selection should act on both components in the same direction. That is, a correlation should exist between the environmental variable and the background component and it should be of the same sign as that between the environmental variable and the frequency of the most active allele. It should be less likely for clines conforming to these expectations to be spurious. Here we have investigated the relationship between the alcohol dehydrogenase locus (*Adh*) polymorphism and temperature in Spanish populations of *Drosophila melanogaster* and have found that it meets the above criterion.

It has been shown that the specific activity is higher for the *Adh^F* than for the *Adh^S* allelic variant of the enzyme⁷⁻⁹, and that loci other than *Adh* affect ADH activity^{4,5}, although the extent of the two effects has been questioned¹⁰. A temperature cline has been reported for this locus⁸, but a widespread association of the *Adh^S* allele with an inversion in the same chromosome arm (II left; refs 11, 12 and our unpublished results) makes interpretation of the cline difficult.

We have chosen the average temperature of the coldest month of the year ($t_{\min}$) as the environmental variable. (Average temperature of the year ($\bar{t}$) and average temperature of the hottest month ($t_{\max}$) are correlated with it over locations; $r_{t_{\min}, \bar{t}} = 0.961$, d.f. = 22, $P < 0.001$; $r_{t_{\min}, t_{\max}} = 0.712$; d.f. = 22,

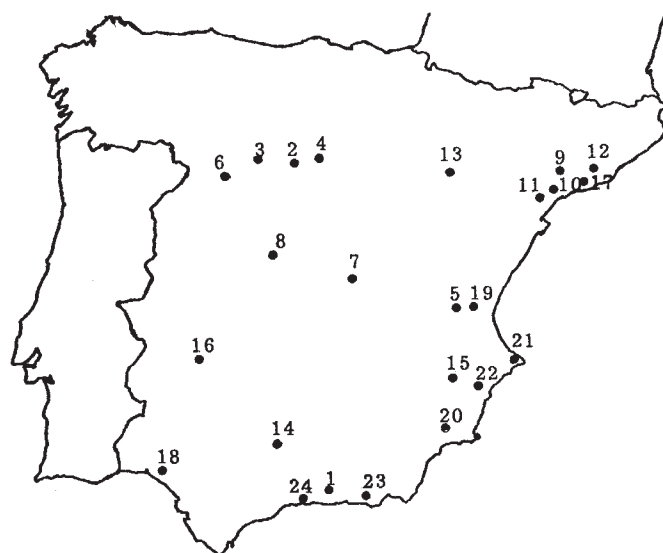


Fig. 1 Geographical location of capture points. Numbers correspond to those assigned to the samples in Table 1.

Table 1 Temperature clines of *Adh* and ADH activity measured in *Adh^{F/F}* females

Sample	Minimum temperature	<i>Adh^F</i> frequency*		Enzyme activity of <i>Adh^{F/F}</i> individuals†		Arrangements <i>Adh^F</i> chromosome‡	
		<i>n</i>	Frequency	No. of groups of 10 ♀♀	per fly	Standard	LAI
1 Alpujarra	-0.31	119	0.962				
2 Peñafiel§	1.29	96	0.948			45	1
3 I Cigales	2.10	94	0.957	9	3.118		
II§	2.10	95	0.947			47	0
4 Fuentespina	2.35	104	0.899	16	3.084		
5 Utiel	2.35	129	0.950	11	2.937		
6 I Toro	2.86	121	0.992	12	2.843		
II§	2.86	96	0.995			70	2
7 Fuente P.N.	3.34	124	0.952	2	2.250		
8 Cebreros§	4.07	95	0.984			64	2
9 Montblanch	4.08	126	0.917	15	2.928		
10 Marsá	4.93	121	0.934	15	3.201		
11 Gandesa	5.62	121	0.955				
12 S. Quintín	5.79	124	0.948	16	2.579		
13 Cariñena	5.95	73	0.932	15	2.073		
14 I Baena	6.51	76	0.882				
II§	6.51	118	0.915			43	0
15 Jumilla	7.27	123	0.866	11	2.885		
16 Castuera	7.36	122	0.906	3	2.420		
17 Vendrell	8.41	108	0.944	15	1.989		
18 Almonte	8.78	121	0.756				
19 Cheste	9.32	126	0.849	7	2.276		
20 Alhama§	9.40	95	0.653			49	0
21 Jaldón	9.47	148	0.730	7	1.814		
22 I Aspe	10.07	127	0.728	15	1.780		
II§	10.07	95	0.742			82	3
23 Egido§	11.81	96	0.745			42	1
24 Torre de Mar	13.63	53	0.877				
Correlation with <i>t_{min}</i>			-0.749		-0.764¶		

* Results from captured wild flies; no significant differences were observed between sexes.

† ADH activity was assayed in descendants of the captured individuals, kept in population cages for no more than 3 months. The heads were electrophoresed and the bodies of *Adh^{F/F}* individuals pooled in groups of 10. Reaction mixture: 0.5 ml of 6 mM NAD, 1 M isopropanol + 0.1 ml of homogenate (10 flies per ml), all in 0.05 M Tris-HCl buffer, pH = 9.0. Assay temperature 25°C. Results in ΔE per m per fly, at 340 nm.

‡ Chromosomes extracted from wild captured males and assayed for cross-over suppression against an *all* chromosome¹⁵. LAI, Left arm inversion.

§ Samples captured in the same month, but a year later than the rest.

|| Angular transformation has been applied to the frequencies (d.f. = 26, $P < 0.001$).

¶ d.f. = 13, $P < 0.001$.

$P < 0.001$.) There is a close correlation between the angular transform of *Adh^F* frequency and *t_{min}* (Table 1). The cline is not strictly latitudinal, in contrast with that observed on the US East Coast⁸ (see Fig. 1 and Table 1).

We have assumed the among-populations differences in ADH activity, measured in *Adh^{F/F}* individuals, to be due to the genetic background. This is a valid approach as the available evidence shows that there is little structural variability segregating within *Adh* electromorphs¹³. ADH activity was measured in *Adh^{F/F}* individuals from 15 locations ($r_{\text{ang.transf. Adh}^F \text{ freq., } t_{\text{min}}} = -0.747$) and a negative correlation with *t_{min}* was found (Table 1). Activities were expressed on a per fly basis as no significant over-locations correlations were found between body weight (x_1) and *t_{min}* or between body weight and activity per fly (x_2) ($r_{x_1, t_{\text{min}}} = -0.2814$, d.f. = 10, $0.4 > P > 0.3$; $r_{x_1, x_2} = 0.1339$, d.f. = 11, $0.7 > P > 0.6$).

In general, this type of result should reduce the probability of a non-selective origin for an observed environmentally correlated allozyme pattern. In our particular case, we note that the environmental variable is an *a priori* one, in the sense that a similar cline has been described already⁸. The other potential source of spuriousness, non-random association, can be reasonably excluded, as the non-random associated gene (or block of them) would have to fulfill the following conditions: (1) to be selected, (2) to have a relevant role in determining the enzymatic activity, and (3) to be associated by chance with the electrophoretic allele that modifies the activity in the same direction. In the present case, as the above stated association between *Adh^S* and the inversion is not absolute (our unpublished results), we have checked the frequencies of left arm inversions within *Adh^F* chromosomes for some of the samples in the cline (Table 1). They were always small, and no significant differences were found between samples located in the first and

the second half of the temperature range ($\chi^2_1 = 0.335$, $0.6 > P > 0.5$).

We regard these results as strong support for the hypothesis of a selective origin of the *Adh* cline in Spanish populations of *D. melanogaster*. We do not imply that temperature is the agent of the probable selection in the cline; any other temperature-correlated variable (such as average concentration of alcohol in the food available to the flies) or a combination of them could be the actual agent.

The present study provides a new approach for investigating the selective origin of clines. We believe it could be applied to a fair proportion of enzymatic polymorphisms as, in most cases, there are differences in the specific activities between allelic forms¹, and, at least in some species, spatial patterns of gene frequencies are not uncommon^{11,14}.

L. Silió carried out the activity assays. We thank F. García-Olmedo, G. N. Somero and C. Wills for their help in the preparation of this manuscript.

Received 25 February; accepted 15 May 1980.

- Harris, H. *Can. J. Genet. Cytol.* **13**, 381-396 (1971).
- Clarke, B. *Genetics* **79**, 101-113 (1975).
- Lewontin, R. C. *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York, 1974).
- Ward, R. D. *Genet. Res.* **26**, 81-93 (1975).
- McDonald, J. F. & Ayala, F. *Genetics* **89**, 371-388 (1978).
- Bulfield, G., Moore, E. A. & Kacser, H. *Genetics* **89**, 551-561 (1978).
- Gibson, J. *Nature* **227**, 959-960 (1970).
- Vigue, C. L. & Johnson, F. M. *Biochem. Genet.* **9**, 213-227 (1973).
- Day, T. H., Hillier, P. C. & Clarke, B. *Biochem. Genet.* **11**, 155-165 (1974).
- Clarke, B., Camfield, R. G., Galvin, A. M. & Pitts, C. R. *Nature* **280**, 517-518 (1979).
- Voelker, R. A. *et al. Genetics* **88**, 515-527 (1978).
- Briscoe, D. A., Robertson, A. & Malpica, J. M. *Nature* **255**, 148-149 (1975).
- Sampsel, B. *Biochem. Genet.* **15**, 971-988 (1977).
- Powers, D. A. & Place, A. R. *Biochem. Genet.* **16**, 593-607 (1978).
- Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Inst. Wash. Publ. 627, 1968).

Feline sarcoma virus polyprotein P115 binds a host phosphoprotein in transformed cells

Fred H. Reynolds Jr & Wim J. M. Van de Ven

Carcinogenesis Intramural Research Program, Frederick Cancer Research Center, Frederick, Maryland 21701

John R. Stephenson

Laboratory of Cellular and Molecular Biology, National Cancer Institute, Frederick, Maryland 21701

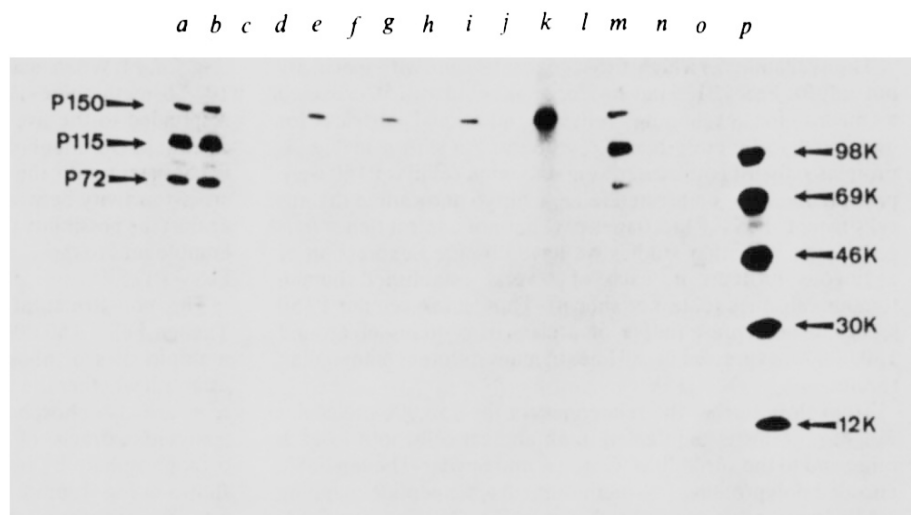
Several independent isolates of feline sarcoma virus (FeSV) have been described¹⁻³. Such viruses are apparently derived by genetic recombination between feline leukaemia virus (FeLV) genomic RNA and host cellular genetic sequences with transforming potential⁴. Two FeSV isolates, one originally described by Gardner and the second by Snyder-Theilen, have been shown to encode polyproteins of around 115,000 molecular weight^{5-8,18,19}. Both polyproteins contain FeLV structural components (p15, p12) at their amino terminus in addition to nonstructural carboxyl terminal components encoded by acquired sequences within the FeSV genome. We have previously shown that Gardner FeSV P115 contains multiple sites of phosphorylation within its nonstructural component and possesses an associated protein kinase activity⁹. In the present study we describe the expression in cells derived from a number of mammalian species, of a highly conserved cellular phosphoprotein with binding affinity for Gardner FeSV P115. This protein, designated P150, exhibits an associated protein kinase activity and is immunologically and structurally distinct from polyproteins encoded by the Gardner or Snyder-Theilen strains of FeSV.

In view of accumulating evidence implicating the non-structural components of Gardner and Snyder-Theilen FeSV encoded polyproteins in transformation (see ref. 10 for review), we analysed uninfected and FeSV transformed mammalian cells

for the presence of proteins which exhibit specific binding affinity, and thus possible functional interaction with FeSV P115. Initially, ¹⁴C-amino acid labelled control and Gardner FeSV transformed mink cells were analysed by immunoprecipitation and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using goat antisera directed against detergent-disrupted FeLV. In addition, a high-titred antiserum prepared by immunization of a goat with FeSV P115 partially-purified from Gardner FeSV(4070-A) pseudotype virions by gel filtration was used. As shown in Fig. 1, immunoprecipitation of FeSV P115 and its major cleavage product, P72, from Gardner FeSV transformed mink cells was obtained with either anti-FeSV P115 (lane b), or anti-FeLV (lane a), but not by normal goat serum (lane c). The viral specificity of these proteins is indicated by their absence from control mink cells (lanes d-f). Precipitation of Gardner FeSV P115 by anti-FeLV has previously been shown to be due to immunological recognition of FeLV p15 and p12 structural components^{6,8,18,19}. An additional protein, of around 150,000 molecular weight (P150), was precipitated from both FeSV transformed and control mink cells by anti-FeSV P115 and from FeSV transformed but not control mink cells by anti-FeLV. These findings demonstrate recognition of P150 by our anti-Gardner FeSV P115 serum and raise the possibility that FeSV P115 may have binding affinity for P150. Such affinity could account for immunoprecipitation by anti-FeLV of P150 from FeSV transformed but not control mink cells.

To explain the above results it is necessary to account for precipitation of P150 from control mink cells by anti-FeSV P115. One possibility is that P150 might have been incorporated into FeSV pseudotype virions as a consequence of an affinity for FeSV P115. Antisera prepared by immunization with partially-purified FeSV P115 might thus recognize mink cellular P150 due to low level P150 contamination of the original immunogen. Alternatively, immunoprecipitation of P150 by anti-FeSV could reflect immunological cross-reactivity between the two proteins. A high-titred antiserum was prepared by repeated immunization of a Fisher rat with FeSV nonproductively-transformed syngeneic tumour cells. If P150 and FeSV P115 shared immunological determinants, sera prepared by this means should also recognize mink cellular P150. As shown in Fig. 1, immunoprecipitation of FeSV P115 and P72 from FeSV

Fig. 1 Analysis by SDS-PAGE of polypeptides encoded by the Gardner strain of FeSV and 150,000 molecular weight cellular phosphoproteins. Cell lines have been previously described^{6,8} and include a fetal mink cell line, CCL64 (d-f, k, l, n); a Gardner FeSV transformed subclone of CCL64, designated 64F3 (a-c, m), a kitten embryonic lung cell line, KeLu (g, h, o) and a canine embryo thymus line, FCf2Th (i, j). Metabolic labelling of exponentially growing cells for 2 h at 37 °C with either ¹⁴C-amino acids (10 µCi per ml) (57 µCi/m Atom, Amersham) (a-j; m-o) or ³²P-orthophosphate (100 µCi per ml) (carrier-free, NEN) (k, l) was performed as reported previously⁹. Following labelling, cell monolayers were washed twice in serum-free Dulbecco's modified Eagles medium and disrupted at 4 °C in 10 mM sodium phosphate pH 7.2, 0.9% NaCl, 1% Triton X-100, 0.5% sodium deoxycholate and 0.1% SDS (PBSTDS) lysis buffer. Extracts (5 ml) were incubated at 4 °C for 18 h following addition of RNase A (200 µg ml⁻¹), RNase T₂ (0.1 µg ml⁻¹), DNase I (4.0 µg ml⁻¹), normal goat serum (25 µl), protein-A Sepharose Cl-4B (1.0 ml of 10% v/v suspension) (Pharmacia), and clarified by centrifugation at 100,000 g for 1 h. Specific proteins were immunoprecipitated by addition of 5 µl of specific antisera to 1 ml lysate, incubation for 18 h at 4 °C, addition of 50 µl of a 10% (v/v) protein-A Sepharose Cl-4B suspension and incubation for a further 2 h at 4 °C. Immunocomplexes were collected by centrifugation at 2,000g for 10 min, washed three times in PBSTDS, resuspended in 20 µl 0.65 M Tris-HCl pH 6.7, 1% SDS, 10% glycerol, 2.5% 2-mercaptoethanol and 0.1% bromophenol blue and heated for 2 min at 90 °C. Finally, samples were electrophoresed through a 5% acrylamide stacking gel and a 9 cm 5-20% linear polyacrylamide (30:0.8 acrylamide: bis-acrylamide) gradient separation gel in a Tris-glycine-SDS-buffer system at 20-30 mA (ref. 17). Radioactivity was visualized by autoradiography using Kodak X-Omat XR2 film. Sera used included goat anti-FeLV (a, d); goat anti-FeSV P115 (b, e, g, i, k), normal goat serum (c, f, h, j, l) and serum from a Fisher rat repeatedly immunized with syngeneic FeSV transformed tumour cells (m-o). Molecular weight standards include ¹⁴C-labelled phosphorylase B (98,000), bovine serum albumin (69,000), ovalbumin (46,000), carbonic anhydrase (30,000) and cytochrome c (12,000) (P).



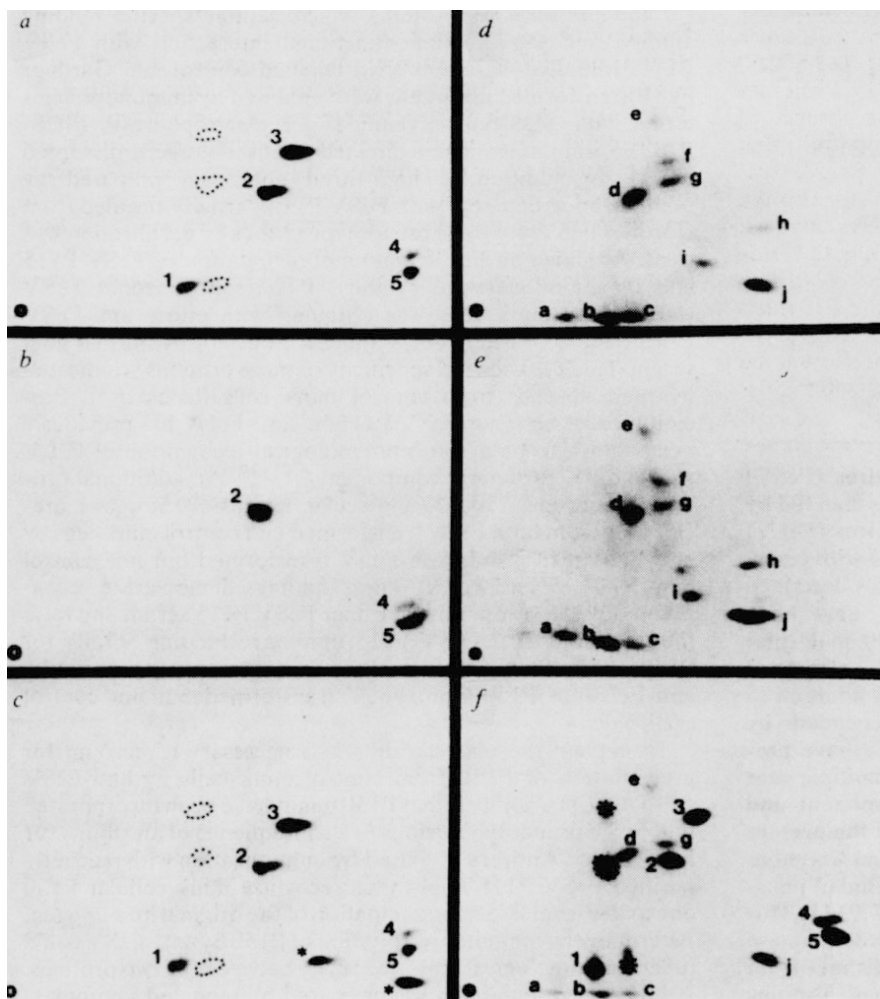


Fig. 2 Comparison by two dimensional tryptic peptide analysis of Gardner and Snyder-Theilen FeSV P115s to the 150,000 molecular weight phosphoproteins expressed in mink and cat cells. Proteins were labelled for 2 h at 37 °C with ^{35}S -methionine (200 μCi per ml) (1075 Ci mmol^{-1} , Amersham) and purified by immunoprecipitation and SDS-PAGE as described in the legend to Fig. 1. Following electrophoresis, gels were washed extensively with 10% trichloroacetic acid, and subsequently with a 10% solution of acetic acid and methanol in water and dried under vacuum. Labelled proteins were localized by autoradiography, and selected gel slices further washed in methanol (10%) and lyophilized. Proteins were digested by incubation of gel slices in tolylsulphonyl phenylalanyl chloromethyl ketone (TPCK)-trypsin (Worthington Biochemicals) (50 μg ml^{-1} in 0.05 M ammonium bicarbonate pH 8.0) for 6 h at 37 °C. Following filtration through a 0.45 μm Millex filter (Millipore) and removal of ammonium bicarbonate by repeated lyophilization, tryptic digests were incubated for 2 h at 0 °C in 0.1 ml of chilled performic acid (30% H_2O_2 and 90% formic acid (1:9) preincubated for 2 h at room temperature). Samples were diluted with 2 ml H_2O , lyophilized, resuspended in electrophoresis buffer comprised of acetic acid:formic acid:water (15:5:80) and spotted on 10 $\times$ 10 cm cellulose coated TLC glass plates (EM Laboratories). Electrophoresis was performed for 30 min at 500 V and chromatography was conducted in the second dimension in buffer containing butanol, pyridine, acetic acid and water (32:25:5:20). Radiolabelled tryptic peptides were visualized by autoradiography using Kodak X-Omat XT 2 film. Detection of peptides was enhanced by spraying TLC plates with a solution of 7% diphenyloxazole in diethyl ether. Proteins analysed included: Gardner FeSV P115, (a); Snyder-Theilen FeSV P115 (b); a mixture of FeLV Pr65^{ra*} (*) and Gardner FeSV P115 (1-5) (c); CCL64 P150 (d); KeLU P150 (e); and a mixture of KeLU P150 (a-j) and Gardner FeSV P115 (f).

transformed (lane m), but not control mink (lane n) or cat (lane o) cells was obtained. The absence of detectable precipitation of P150 from uninfected control cells (lane n) argues against the possibility that P150 and FeSV P115 possess common immunological determinants.

The availability of a high-titred goat antiserum with specificity not only for FeSV P115 but also for mink cellular P150 provided a means for examining cells of additional species for immunologically cross-reactive proteins. As shown in Fig. 1, proteins similar in molecular weight to mink cellular P150 were precipitable from uninfected feline embryo and canine thymus cells by anti-FeSV P115 (lanes g, i) but not control (lanes h, j) goat serum. In other studies we have observed expression of analogous proteins in each of several established human tumour cell lines (data not shown). Thus, mink cellular P150 seems to be representative of a class of proteins of around 150,000 M , expressed in cell lines of many different mammalian species.

To explore further the relatedness of the 150,000 molecular weight proteins in uninfected mink and cat cells both to each other and to the 115,000 M , Gardner and Snyder-Theilen FeSV encoded polypeptides, ^{35}S -methionine tryptic peptide mapping studies were performed. As shown in Fig. 2a, Gardner FeSV P115 contains five ^{35}S -methionine labelled peptides, none of which correspond to the two ^{35}S -methionine labelled peptides specific to the p30 component of FeLV Pr65^{ra*} (Fig. 2c). Three of the five Gardner FeSV P115 peptides are also represented in Snyder-Theilen FeSV P115 (Fig. 2b) consistent with the contention that these independently-derived FeSV isolates possess common or overlapping acquired genetic sequences^{4,8,19}. Additional minor peptides within the Gardner FeSV P115 peptide map, indicated by dotted circles, are probably

nonspecific in that their presence is variable, and in contrast to the five major ^{35}S -methionine labelled peptides, is dependent on the species of origin of cells analysed. Analysis of mink and cat cellular P150s revealed 10 major ^{35}S -methionine labelled peptides which seemed very similar by two dimensional analysis (Fig. 2d, e). When analysed in a mixing experiment, none of the 10 ^{35}S -methionine labelled peptides specific to P150 corresponded to the five major ^{35}S -methionine labelled peptides present in Gardner FeSV P115 (Fig. 2f). This latter observation is consistent with the above described lack of immunological cross-reactivity between these two proteins and further argues against the possibility that P150 is encoded by cellular sequences homologous to those encoding the non-structural component of FeSV P115.

The non-structural components of Gardner and Snyder-Theilen FeSV 150,000 M , polypeptides are known to contain multiple sites of phosphorylation⁹. It was thus of interest to establish whether the above described 150,000 M , cellular proteins are also phosphorylated. To examine this possibility, we analysed extracts of cells metabolically labelled with ^{32}P -orthophosphate by immunoprecipitation and SDS-PAGE. As shown in Fig. 1, mink cellular P150 was efficiently labelled and was immunoprecipitated from mink cells by anti-FeSV P115 (lane k) but not normal goat serum (lane l), establishing its phosphorylated nature. Tryptic peptide analysis of ^{32}P -labelled P150s from mink and cat cells revealed each to possess 10 to 11 major well-resolved phosphopeptides. The high degree of relatedness of the tryptic peptide patterns of the two proteins further indicates that these proteins are relatively conserved (data not shown).

We have previously reported⁹ the presence of readily detectable levels of protein kinase activity in immunoprecipitates of

Gardner FeSV P115. The polyprotein itself has been shown to represent a major substrate for this enzymatic activity. Because of the above findings indicating the invariable presence of P150 in such immunoprecipitates, it was of interest to find whether cellular P150s exhibited similar reactivities in the absence of FeSV P115. Extracts of FeSV transformed and control mink cells were prepared, immune precipitated by anti-FeLV, anti-FeSV P115 and normal goat serum and analysed for protein kinase activity. Enzymatic reactivities and substrates involved were identified by SDS-PAGE. As shown in Fig. 3, both Gardner FeSV P115 and mink cellular P150 were efficiently labelled by this means. Incorporation of phosphate into both P150 and FeSV P115 was in the range of 10–20 fmol per min. Even in the absence of FeSV P115, efficient labelling of P150 was observed (lane *e*) establishing that *in vitro* [γ - 32 P]ATP labelling of P150 is not dependent on an FeSV P115 associated protein kinase. Additional substrates for the P150 associated protein kinase included proteins of around 45,000–50,000 molecular weight in control and FeSV transformed mink cells and a 40,000 molecular weight protein detectable only in control mink cells. The significance of these proteins and their relationship to P150 remain to be determined.

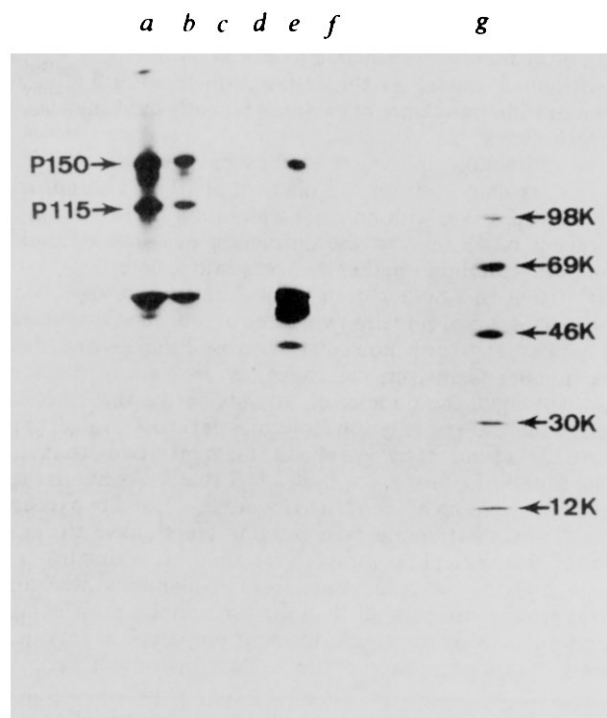


Fig. 3 Analysis of Gardner FeSV P115 and mink cellular P150 containing immunoprecipitates for protein kinase activity. FeSV transformed (*a–c*) and control (*d–f*) CCL64 mink cells were disrupted by repeated aspiration of 1.0×10^7 cells in 5 ml 10 mM Tris-HCl pH 7.2, 100 mM NaCl, 1 mM EDTA and 0.1% Triton X-100 through a 25-gauge needle. Disrupted cells were clarified by centrifugation at 2,000 r.p.m. for 10 min, preabsorbed by addition of 50 μ l normal goat serum and 0.5 ml of a 10% (v/v) suspension of protein-A Sepharose CL-4B (Pharmacia), incubated overnight at 4°C and reclarified by centrifugation at 100,000g for 2 h. Following addition of 5 μ l specific antiserum, lysates (1 ml) were incubated at 4°C for 18 h, 50 μ l of a 10% (v/v) protein-A Sepharose suspension was added and incubation continued for a further 2 h at 4°C. Immunocomplexes were collected by centrifugation for 10 min at 2,000g and washed in PBSTDS. Protein kinase activity was assayed by resuspension of immunoprecipitates in 20 mM Tris-HCl, pH 7.2, 5 mM $MgCl_2$ buffer containing 1 μ Ci [γ - 32 P]ATP (2000 Ci mmol $^{-1}$, NEN) and incubation for 10 min at 30°C. Reactions were terminated by addition of 3 ml ice cold PBSTDS and nonprotein-bound [32 P] label removed by repeated centrifugation in excess PBSTDS. After the final wash, immunoprecipitates were analysed by SDS-PAGE as described in the legend to Fig. 1. Sera used included goat anti-FeLV (*a, d*); goat anti-FeSV P115 (*b, e*) and normal goat (*c, f*). Molecular weight standards are as described in the legend to Fig. 1.

Immunoprecipitation of P150 from both normal and FeSV transformed cells of various species by the anti-FeSV P115 serum used in the present study suggests that P150 is incorporated into pseudotype virions in association with FeSV P115 and represents a probable minor contaminant of the purified FeSV P115 used as the original immunogen. To test this possibility directly, immune precipitates of density-gradient purified FeSV(4070-A) and control 4070-A virions were analysed for protein kinase activity. As shown in Fig. 4, both P150 and FeSV P115 were precipitated from FeSV (4070-A) pseudotype virions by anti-FeLV but not by normal goat serum.

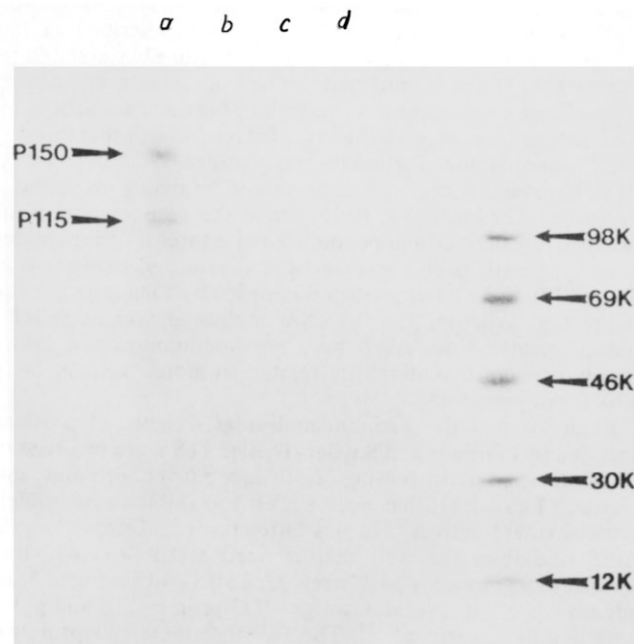


Fig. 4 Association of P150 and FeSV P115 in FeSV(4070-A) pseudotype virions. FeSV(4070-A) (*a, b*) and 4070-A (*c, d*) virions were purified by sucrose density gradient centrifugation, disrupted and analysed for protein kinase activity as described in Fig. 3. Sera used included goat anti-FeLV p12 (*a, c*) and normal goat (*b, d*). Molecular weight standards are as described in the legend to Fig. 1.

Both proteins represented efficiently labelled substrates for protein kinase activities present in these immune precipitates. Proteins of around 40,000–50,000 *M*, seen in immunoprecipitates of FeSV transformed and control mink cells were not present in immunoprecipitates prepared from FeSV pseudotype virus, arguing against the possibility that their presence is necessary for the protein kinase activities associated with FeSV P115 and P150.

The present findings establish the expression in mink cells of a highly conserved 150,000 molecular weight cellular phosphoprotein with binding affinity for Gardner FeSV P115. Although the biological role of this cellular phosphoprotein is unknown, its affinity for FeSV P115 may have functional significance. P150 and FeSV P115 both represent efficient substrates for protein kinase activity present in immunoprecipitates of FeSV P115. While *in vitro* [γ - 32 P]ATP labelling of mink cellular P150 is shown to proceed in the absence of FeSV P115, the converse is not necessarily true. Thus, the possibility that P150 possesses intrinsic protein kinase activity and is responsible for phosphorylation of FeSV P115 must be considered. Such a model would be consistent with the affinity that these proteins exhibit for each other. The possibility that both P150 and FeSV P115 possess intrinsic protein kinase activity or alternatively, that an as yet undetected cellular protein kinase is responsible for phosphorylation of one or more of these proteins cannot, however, be excluded.

It is of interest to compare the 150,000 molecular weight cellular phosphoproteins described above to a protein of similar molecular weight recently shown to be expressed in mouse thymus and to a lesser extent, spleen and bone marrow cells¹¹. This latter protein, designated NCP 150, has not as yet been characterized with respect to phosphorylation and protein kinase activity. However, NCP 150 has been shown to share immunological determinants with the non-structural component of a 120,000 molecular weight polypeptide encoded by Abelson murine leukaemia virus and is apparently encoded by normal cellular sequences analogous to acquired cellular sequences represented within the AbLV genome. In contrast, on the basis of immunological and structural criteria, the 150,000 molecular weight cellular protein described in the present study is immunologically and structurally unrelated to FeSV P115. The possibility that this lack of relatedness is due to genetic drift subsequent to acquisition of cellular sequences by the FeSV genome seems to be excluded on the basis that three of five ³⁵S-methionine labelled tryptic peptides of Gardner FeSV P115 have been found to be represented by analogous peptides in Snyder-Theilen FeSV P115. Since the two sarcoma virus isolates were derived independently, these three tryptic peptides have apparently been conserved and should thus be present in feline P150 if the latter protein is encoded by analogous genetic sequences. Whether the 150,000 molecular weight cellular phosphoproteins described here are immunologically, structurally or even evolutionarily related to mouse cellular NCP 150, is not yet clear.

Evidence that the 115,000 molecular weight polypeptides encoded by Gardner and Snyder-Theilen FeSV are involved in transformation is increasing. Both are phosphoproteins and Gardner FeSV P115 has been shown to exhibit an associated protein kinase activity⁹. In this latter respect, Gardner FeSV P115 resembles the well characterized avian sarcoma virus transforming protein, pp60^{src} (refs 12, 13), SV40 T antigen¹⁴, an adenovirus-SV40 hybrid T antigen (D2 protein)^{15,16} and polyoma middle T antigen²⁰⁻²². The fact that these polypeptides, encoded by independently-acquired cellular sequences, share tryptic peptides, argues for the existence within the normal cellular genome of a limited number of genetic sequences which, on insertion into the type C viral genome, can lead to malignant transformation. The identification of a highly conserved cellular phosphoprotein with associated protein kinase activity and binding affinity for FeSV P115 may have significance with respect to the elucidation of cellular regulation mechanisms involved in transformation of eukaryotic cells.

We thank S. E. VanLaningham-Miller and R. P. Nalewaik for technical assistance. This work was supported under NCI contract NOI-CO-75380, and by a fellowship awarded to W. J. M. Van de Ven by The Netherlands Organization for the Advancement of Pure Research (ZWO).

Note added in proof: Gardner FeSV P115-associated protein kinase specifically phosphorylates tyrosine acceptor sites represented both within P115 itself and in cellular P150.

Received 2 April; accepted 2 May 1980.

- Snyder, S. P. & Theilen, G. H. *Nature* **221**, 1074-1075 (1969).
- Gardner, M. B. *et al. Nature* **226**, 807-809 (1970).
- McDonough, S. K., Larsen, S., Brodey, R. S., Stock, N. D. & Hardy, W. D. *Jr Cancer Res.* **31**, 953-956 (1971).
- Frankel, A. E., Gilbert, J. H., Porzig, K. J., Scolnick, E. M. & Aaronson, S. A. *J. Virol.* **30**, 821-827 (1979).
- Khan, A. S. & Stephenson, J. R. *J. Virol.* **23**, 599-607 (1977).
- Stephenson, J. R., Khan, A. S., Sliski, A. H. & Essex, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5608-5612 (1977).
- Sherr, C. J., Sen, A., Todaro, G. J., Sliski, A. & Essex, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1505-1509 (1978).
- Van de Ven, W. J. M., Khan, A. S., Reynolds, F. H. Jr, Mason, K. & Stephenson, J. R. *J. Virol.* **33**, 1034-1045 (1980).
- Van de Ven, W. J. M., Reynolds, F. H. Jr & Stephenson, J. R. *Virology* **101**, 185-197 (1980).
- Stephenson, J. R., Khan, A. S., Van de Ven, W. J. M. & Reynolds, F. H. Jr *J. natn. Cancer Inst.* **63**, 1111-1119 (1979).
- Witte, O. N., Rosenberg, N. E. & Baltimore, D. *Nature* **281**, 396-398 (1979).
- Collett, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021-2024 (1978).
- Levinson, A. D., Oppermann, H., Levintow, L., Varmus, H. E. & Bishop, J. M. *Cell* **15**, 561-572 (1978).
- Griffin, J. D., Spangler, G. & Livingston, D. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2610-2614 (1979).

- Baumann, E. A. & Hand, R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3688-3692 (1979).
- Tjian, R. & Robbins, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 610-614 (1979).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Khan, A. S., Deobagkar, D. N. & Stephenson, J. R. *J. biol. Chem.* **253**, 8894-8901 (1978).
- Barbacid, M., Lauver, A. V. & Devare, S. G. *J. Virol.* **33**, 196-207 (1980).
- Griffin, J. D., Spangler, G. & Livingston, D. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2610-2614 (1979).
- Schaffhausen, B. S. & Benjamin, T. L. *Cell* **18**, 925-933 (1979).
- Smith, A. E., Smith, R., Griffin, B. & Fried, M. *Cell* **18**, 915-924 (1979).

Fraying of A-filaments into three subfilaments

Maria C. Maw & Arthur J. Rowe

Department of Biochemistry, School of Biological Sciences, University of Leicester, Leicester LE1 7RH, UK

Since Huxley¹ established their bipolar structure, direct electron microscopy of A-filaments isolated from vertebrate skeletal muscle has yielded little further interpretable information about the mode of packing of the myosin molecules within the filament. Using A-filaments prepared from rat psoas muscle we have now found it possible to induce clear fraying of these filaments into subfilaments, by exposure of the preparation to very low ionic strength before contrasting with uranyl acetate. The number of such filaments observed is generally (and never in excess of) three. Considerations of symmetry suggest that the formation of these frayed filaments is compatible only with a 'three-stranded' model for the native A-filament, a finding in agreement with the balance of evidence recently published using other techniques²⁻⁵.

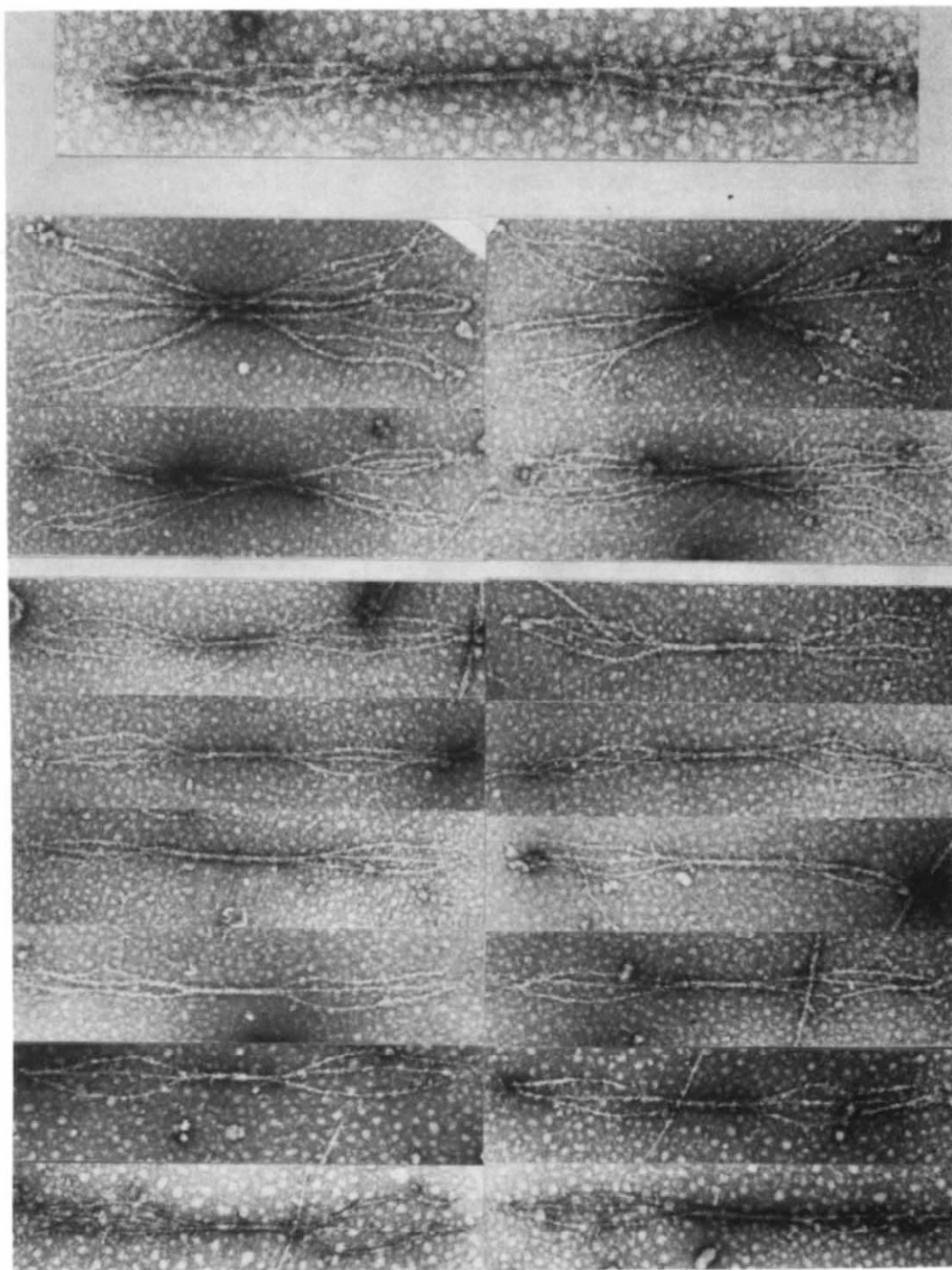
X-ray diffraction and other studies have shown that the cross-bridges emerge from the filament shaft in an approximately helical array, with an axial translation of 14.3 nm and axial repeat of 43 nm, but the diffraction evidence does not indicate with certainty whether the cross-bridge helix is 2-, 3- or 4-start^{6,7}: that is, whether each 'crown' of cross-bridges (one 'crown' per 14.3 nm) contains two, three or four such structures. If the number of myosin molecules per cross-bridge is one, then n (the number of myosin molecules per 14.3 nm of filament length) will equal the number of 'strands' in the three helical models. If there are two myosin molecules per cross-bridge, then it is virtually certain that $n = 4$, and the filament is two-stranded. Estimates published for n are shown in Table 1. As Squire has pointed out², opinion has tended to favour $n = 3$, and two recent studies^{3,4} which determine mass per unit length have strongly supported this value, in contrast to certain other estimates⁵.

While studying the reconstruction of A-filaments⁸ we have used rat psoas muscle rather than the usual rabbit muscle, and electron microscopy of relaxed filament preparations revealed that a significant proportion of the A-filaments were 'frayed' at

Table 1 Estimates by various workers for n , the number of myosin molecules per 14.3-nm repeat in the A-filament of vertebrate skeletal muscle

Method	Estimated value for n	Ref.
Inference from observed symmetry (E/M)	3	This work
Quantitative SDS-PAGE of myofibrils	3.8 ± 0.5	5
	2.5	20
	2.7	21
	3.9	22
Mass per unit length determined by hydrodynamics	3.1 ± 0.2	4
Mass per unit length determined by quantitative electron scattering	2.7 ± 0.7	3
Particle counting (E/M)	4.3	22
Nucleotide binding to myofibrillar protein	3.5 ± 0.2	23
Calculation from myosin content and density of A-filaments per unit volume of muscle	4	24

Fig. 1 Electron micrographs of A-filaments and small a-segments (second and third rows) in relaxed filament preparations prepared from rat psoas muscle by the method of Emes and Rowe⁴. The preparations were rinsed on the support grids first with 0.12 M KCl and then with several drops of distilled water before contrasting with unbuffered 2% uranyl acetate. 'Frayed' filaments of the type shown here are routinely seen when this procedure is followed, both in areas where the filaments are negatively contrasted (as shown) and where positive contrasting has occurred. top, $\times 114,500$; Others, $\times 45,000$.



the ends into subfilaments (Fig. 1). We have occasionally observed this effect with rabbit A-filaments prepared in identical conditions. Filaments that 'fray' are indistinguishable in length or width (of the 'non-frayed' region) from intact filaments (Fig. 2); as the proportion that fray varies from preparation to preparation, and even from area to area within the same specimen grid, it is unlikely to constitute a genuine subpopulation of filaments. The variation presumably arises from exact drying-down conditions which are not yet fully controlled.

More than 800 frayed filaments were photographed, measured and examined. In every case only two or three subfilaments were present. Where two subfilaments were counted one was usually significantly thicker than the other (Fig. 1); the thicker one is presumed to comprise two unfrayed subfilaments. We never observed an A-filament that frayed into more than three subfilaments. Fraying, which commences at the distal region, can proceed to any point along the filament axis

(Fig. 2) up to a central region $\sim 0.3 \mu\text{m}$ wide. Within experimental error this corresponds to the innermost stripe of the A-segment which stains with impure anti-C-protein (that is, stripe 9, following the terminology of Craig and Offer⁹). Although the subfilaments are not smooth and show evidence of projections which may be cross-bridges, a clear repeating structure has not as yet been demonstrated.

Subfilaments formed only when the preparation on the microscope grid was washed with several drops of distilled water before contrasting with uranyl acetate. In control experiments, preparations made by this technique showed from 32 to 79% of frayed filaments in randomly selected areas; however a further wash with 0.12 M KCl before contrasting produces almost complete reversal of the effect. No significant proportion of frayed filaments is found if washing with ammonium formate solution (pH 7, $I = 0.12$) is substituted for the distilled water wash. The simplest explanation for this phenomenon is that at

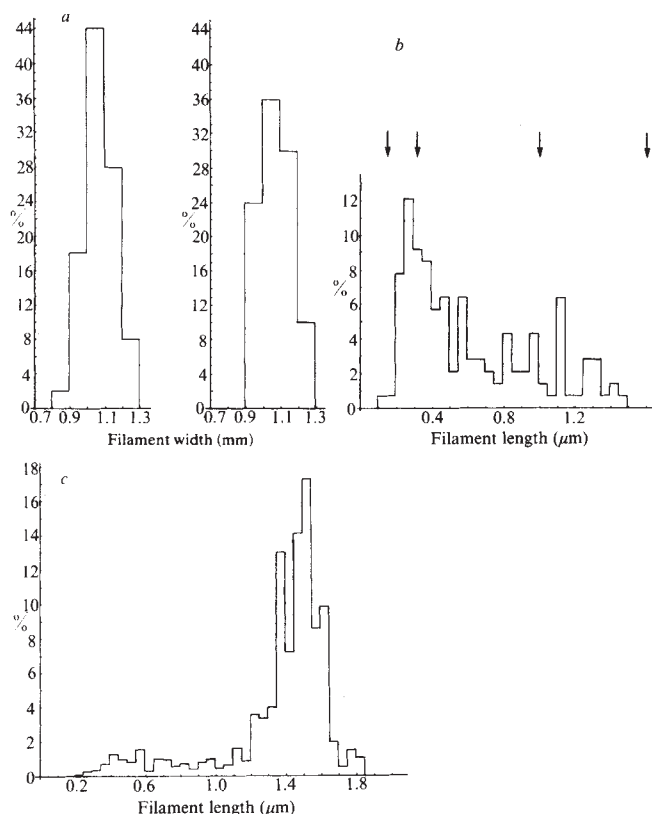


Fig. 2 Statistical analysis of the dimensions of frayed A-filaments showing: *a*, the measured width ($\times 68,700$) of the unfrayed central region, avoiding the presumed M-line region, as compared with the similarly measured width of normal, unfrayed filaments (R) from adjacent fields; *b*, the distribution of the length (from one side of the filament to the other) of the unfrayed central region of the frayed filaments, with vertical arrows showing (from L to R) the distances corresponding respectively to the M-line, stripe 9, stripe 1 (terminology of Craig and Offer⁹) and the filament tip; *c*, the length distribution (weight-averaged) of the frayed filaments, this being completely comparable to that obtained with normal, unfrayed filaments⁴. Approximately 800 filaments measured.

low ionic strength the expansion of the electrical double layer around the myosin molecules leads to greatly increased repulsive forces¹⁰. An increase in net negative charge on the filaments might be expected to have a similar effect, and with synthetic filaments made from rat myosin, Pinset-Härström and Truffly¹¹ show that high concentration of ATP induces a rather poorly defined tendency to fray. The whole phenomenon may be related to one of the first observations made on myosin, by Szent-Györgyi¹², who showed that prolonged dialysis of myosin solutions at low ionic strength (that is, synthetic filament preparations) against distilled water yielded clear but viscous solutions.

We therefore conclude that the A-filaments from rat psoas muscle contain three asymmetric units which appear as subfilaments after fraying. Our micrographs give no clear evidence to suggest that these subfilaments are packed in the intact A-filament into other than a simple linear array: certainly any large degree of coiling of the subfilaments seems implausible in the light of their ability to fray apart and then re-associate whilst loosely adsorbed on to a carbon film. Our evidence does not, of course, yield a direct estimate for *n*, but it is entirely compatible with 3-start helical models supported by recent measurements of mass per unit length of A-filaments^{3,4}.

This finding provides an important constraint on the type of molecular model which can be postulated for the A-filament. An obvious conclusion is that if the tails of the myosin molecules are organized into three subfilaments along the shaft, then all the myosin molecules cannot be equivalent at the level of near-

neighbour contacts in the main cross-bridge region, and may not even be helically packed. This is not, however, a necessary conclusion. If the subfilaments arise as a result of fraying, initiated (as it seems) from the tip region, then detailed model-building (A.J.R., in preparation) shows that a model which will fray into three subfilaments can be devised, which conserves the conventionally made assumptions regarding equivalence and helical packing.

The finding of 3-fold physical symmetry in the A-filament is consistent with data such as the near-triangular appearance of the shaft in cross-sections¹³; however, no currently published model immediately suggests the fraying into three subfilaments now observed. The model of Pepe^{14,15} packs the myosin tails into linear array, but it is two-stranded with respect to the cross-bridge lattice and hence unlikely to fray into three subfilaments. Moreover, recent evidence¹⁶ casts doubt on the existence of the myosin dimer as a kinetic unit, almost certainly an essential feature of a two-stranded, two myosin/cross-bridge model. Squire's model^{7,17}, like that of Wray¹⁸ for crustacean thick filaments, could only fray into subfilaments with pronounced coiling. The finding of Davey and Graafhuis¹⁹ of three highly coiled subfilaments after tryptic digestion has been shown (A.J.R. and M.C.M., in preparation) to be an artefact of trypsin incubation. We conclude that a thorough re-examination of possible models for A-filament structure is now necessary, in which long-held but never fully proved assumptions as to the equivalence and helical packing of the molecular packing units may be called into question.

This work was supported by an MRC grant.

Received 3 March; accepted 30 May 1980.

1. Huxley, H. E. *J. molec. Biol.* **7**, 281–308 (1963).
2. Squire, J. M. *Nature* **281**, 99–100 (1979).
3. Lamvik, M. K. *J. molec. Biol.* **122**, 55–68 (1978).
4. Emes, C. H. & Rowe, A. J. *Biochim. biophys. Acta* **537**, 125–144 (1978).
5. Pepe, F. A. & Drucker, B. J. *J. molec. Biol.* **130**, 379–393 (1979).
6. Huxley, H. E. & Brown, W. J. *J. molec. Biol.* **124**, 383–434 (1967).
7. Squire, J. M. *J. molec. Biol.* **77**, 279–323 (1973).
8. Maw, M. C. & Rowe, A. J. *J. ultrastruct. Res.* **69**, 142–143 (1979).
9. Craig, R. & Offer, G. *Proc. R. Soc. B* **192**, 451–461 (1976).
10. Verwey, E. J. W. & Overbeek, J. Th. G. *The Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
11. Pinset-Härström & Truffly, J. J. *J. molec. Biol.* **134**, 173–188 (1979).
12. Szent-Györgyi, A. *The Chemistry of Muscular Contraction* (Academic, New York, 1951).
13. Luther, P. K. & Squire, J. M. *J. molec. Biol.* **125**, 313–324 (1978).
14. Pepe, F. A. in *Electron Microscopy* Vol. 2 (ed. Uyeda, R.) 53 (Maruzen, Tokyo, 1966).
15. Pepe, F. A. *J. molec. Biol.* **27**, 203–225 (1967).
16. Emes, C. H. & Rowe, A. J. *Biochim. biophys. Acta* **537**, 110–124 (1978).
17. Squire, J. M. A. *Rev. Biophys. Bioengng* **4**, 137–163 (1975).
18. Wray, J. S. *Nature* **277**, 37–40 (1979).
19. Davey, C. L. & Graafhuis, A. E. *Experientia* **32**, 32–34 (1976).
20. Potter, J. D. *Archs Biochim. Biophys.* **162**, 436–441 (1974).
21. Tregear, R. T. & Squire, J. M. *J. molec. Biol.* **77**, 279–290 (1973).
22. Morimoto, K. & Harrington, W. F. *J. molec. Biol.* **83**, 83–97 (1974).
23. Maruyama, K. & Weber, A. *Biochemistry* **11**, 2990–2998 (1972).
24. Huxley, H. E. in *The Cell* Vol. 4 (eds Brachet, J. & Mirsky, A. E.) (Academic, New York, 1960).

The yeast transposon Ty1 generates duplications of target DNA on insertion

Jürg Gafner & Peter Philippsen

Microbiology Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Genetic elements which can transpose into different chromosomal sites causing modulation of gene expression have been found in maize¹ and several other organisms². Most intensively characterized are those in enterobacteria: the insertion sequences, short DNA segments with probably no coding information, the transposable drug-resistance genes and the phage Mu (refs 2–4). The molecular details of transpositions in bacteria are under investigation; important results are the duplication of short target sequences at the integration site^{5–15} and the identification of a transposase protein^{16,17}. The yeast transposable element Ty1 is present in about 35 copies per genome; it is 5,600 nucleotide pairs long with terminal direct

repeats of 300 nucleotide pairs and is able to integrate into new chromosomal sites¹⁸. The resemblance of this yeast element to mobile elements in *Drosophila melanogaster*^{19,20} is striking. The best characterized element, called copia, is present in about 30 copies in the *Drosophila* genome^{19,21} and can integrate into many sites²², in some cases as part of a large transposable segment²³. Another remarkable structural resemblance exists with the proviral DNA of retrovirus, which can insert into different chromosomal sites, as found in avian sarcoma virus transformed rat cells²⁴. We focus here on the DNA sequence at the termini of element Ty1 (ref. 18). Our sequence data show that transposition in yeast and bacteria have several features in common, such as integration into different sites without any apparent sequence specificity and duplication of five nucleotide pairs of target DNA at the integration site. In addition, we find that the terminal repeats in two Ty1 elements consist of identical 338 nucleotide pairs.

The upper part of Fig. 1 shows the basic structure of Ty1 as deduced from heteroduplex experiments with cloned yeast DNA¹⁸; the 300 nucleotide pair repeats, called δ , flank a 5,000-nucleotide transcribed sequence which we call ϵ . Short sequences with homologies to the δ sequences at the Ty1 termini were also found frequently alone in the chromosomal DNA¹⁸. These so-called solo δ sequences show up to 15% sequence divergence as determined by DNA sequence analyses in our early experiments (J. G., unpublished). Determination of whether the δ sequences at the Ty1 termini also show some divergence or whether they are identical should clarify the problem of whether the synthesis of both termini is guided by one template or not. The integrated DNA of retroviruses, for example, probably has identical termini of several hundred nucleotides, as concluded from the known sequence arrangement of the template, the retrovirus RNA³³. On the other hand, the direct or inverted terminal repeats of bacterial transposons need not be identical according to plausible transposition mechanisms^{25,26}. In other words, the two termini need not originate from the same template in the transposition event.

Two separate Ty1 elements were investigated: Ty1-B10, located to the left of the Sup4 locus on chromosome X in strain B596, and Ty1-D15 of unknown but different location in strain D585-11C (ref. 18). Cloned restriction fragments containing the Ty1 termini were sequenced according to the Maxam and Gilbert protocol²⁷ and the strategy for this analysis is outlined in Fig. 1. The reading of the sequence gels reveals identical δ sequences of 338 nucleotide pairs at the four termini. Figure 2 shows part of the autoradiograms for the left terminus of Ty1-D15 and Fig. 3 gives the actual sequences with the identical parts written only once.

A single δ sequence does not show any immediately obvious coding or structural information. It contains 15 start codons and several octanucleotides similar to the sequence 5'-TATAAATA, which are common to eukaryotic promoter

regions²⁸; but the δ sequence also contains 53 stop codons in all reading frames and the longest possible translation product coded by a δ sequence would consist of 42 amino acids. However, there is an open reading frame running out of the δ sequence starting at nucleotides 41-39 in the lower strand (Fig. 3). A search for structural features revealed some short inverted repeats up to 13 nucleotides long (Fig. 3), but these were not located at the ends. Terminal short inverted repeats have been found in bacterial transposable elements and are probably of functional importance^{6,10}.

Work on several other Ty1 elements is in progress. We have seen some divergence from the δ sequence in Fig. 3, but the alterations appear in both ends. It is possible that, despite variations among different Ty1 elements, both termini are always identical.

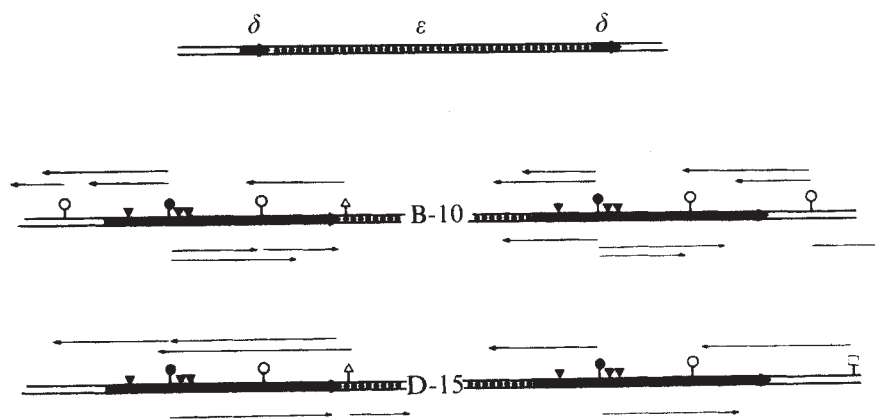
Transposition events in bacteria create duplications of short sequences of target DNA at the integration sites which are then flanking the transposed DNA. DNA sequence analysis shows clearly that the number of the duplicated nucleotides, but not the sequence itself, is specific for a given transposable element: 5 nucleotides for IS2, Tn3 and Mu¹¹⁻¹⁴, 9 nucleotides for IS1, Tn9, Tn10 and Tn903 (refs 5-10) and 11 or 12 nucleotides for IS4 (ref. 15). Analysis of the sequences surrounding the yeast transposable element shows direct repeats of 5 nucleotides flanking the Ty1: 5'-GAAAC in the case of Ty1-B10 and 5'-ATTTT for Ty1-D15 (Fig. 4). For the first case it was demonstrated that these repeats originate from a duplication. As mentioned above, the Ty1-B10 is located to the left of the Sup4 locus on chromosome X. Cloned DNA from this locus of strain S288C, which lacks the Ty1 at this position¹⁸, was analysed. The sequences surrounding the Ty1-B10 were found with the 5'-GAAAC present only once (Fig. 4).

It is unlikely that the pentanucleotide which becomes duplicated is a recognition sequence for Ty1 integration, because this pentanucleotide is different for Ty1-B10 and Ty1-D15. (A third case shows a 5'-GAGAA adjacent to Ty1; P.P., unpublished.) Looking for possible recognition characteristics or preferable sites of Ty1 insertions, we only note that the 30 nucleotides on either side of each Ty1 examined consist of 70% or more A + T pairs (Fig. 4). A preference for A + T-rich regions is also found for the bacterial IS1 transposition²⁹.

Both Ty1 elements analysed here also have direct 5-nucleotide repeats (5'-TACCA) at the junctions between the δ and ϵ segment (Fig. 4). These internal repeats might have functional significance although such internal duplications at comparable locations have not been observed in bacterial transposons.

The solo δ sequences mentioned above probably mark former sites of Ty1 elements which were deleted by homologous recombination between the termini, leaving one δ sequence behind. Such deletion processes are known in bacteria for λ phages carrying the chloramphenicol transposon Tn9, which has direct repeated IS1 termini; the deletion generates λ phage with

Fig. 1 Structure of Ty1 and strategy of DNA sequence analysis. The general structure of an integrated Ty1 is shown in the upper part (chromosomal DNA, double line; δ sequences, thick black arrows; ϵ sequence, dotted line). The δ sequences of Ty1-B10 and Ty1-D15 and of some adjoining DNA were sequenced according to Maxam and Gilbert²⁷. The arrows mark the length and direction of sequences determined in restriction fragments which were labelled at their 3' ends and also sometimes at their 5' ends. *Xho*I (●), *Taq*I (○), *Hinf*I (▼), *Bgl*II (△) and *Hpa*II (□) restriction sites are indicated. Because of the sequence identity of the four δ sequences we did not determine in all four cases the sequence of both DNA strands.



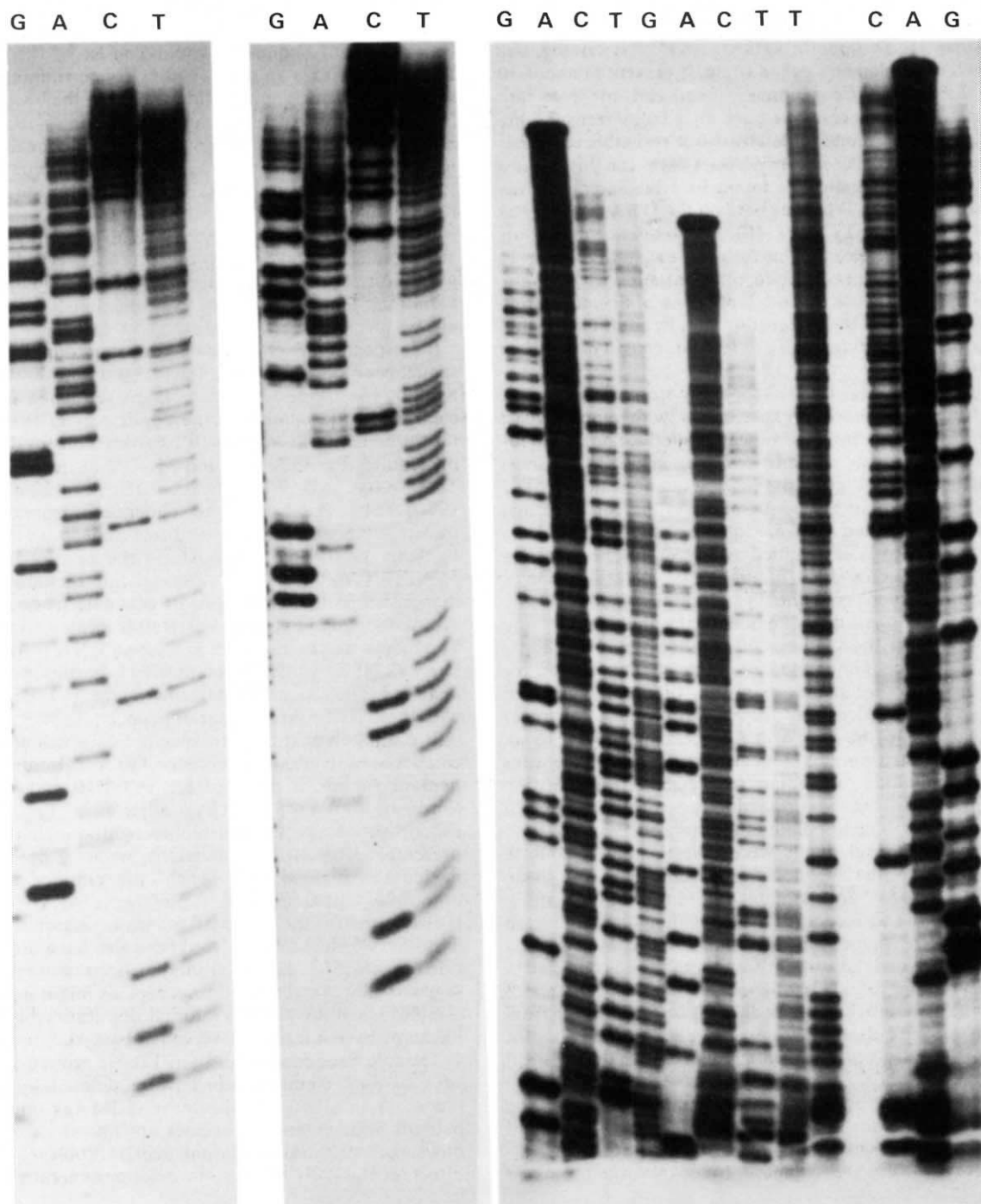


Fig. 2 Autoradiograms of sequence gels (left part of Ty1-D15, Fig. 1). Left, two 18% gels showing the sequence from the *XhoI* site towards the chromosomal DNA and towards the centre of the δ sequence, respectively. Right, 8% gel showing the sequence from the end of the δ sequence towards the *XhoI* site. Electrophoresis of the DNA in the left, middle and right tracks in the 8% gel continued for 8, 12 and 4 h, respectively.

a single *IS1* sequence³⁰. The observed sequence divergence among solo δ sequences (base differences and small deletions) thus probably reflects either random mutations after the Ty1 deletion, the terminal heterogeneity among different Ty1, or both. Clearly, the solo δ sequences are potential targets for a new Ty1 integration if homologous recombination is involved. A comparable process in bacteria is the homologous recombination between two *IS1* sequences from different replicons which creates transposable sequences flanked by two direct repeated *IS1*³¹. An interesting result concerning the question of a similar mechanism for Ty1 insertion comes from the analysis of the target sequence for Ty1-B10.

As presented above, the sequences to left and right of Ty1-B10 are identical to the continuous sequence of the same chromosomal site in strain S288C, which lacks the Ty1-B10, except for the duplicated pentanucleotide. By reading the S288C sequence (upper line, Fig. 4) in the inverse orientation, one recognizes the central part of a δ sequence (nucleotides 237-119 of Fig. 3 with point mutations and a deletion of nucleotides 175-179). By sequencing several hundred additional nucleotides on both sides of the S288C sequence and the Ty1-B10, we confirmed the Ty1-B10 transposition into the middle of a solo δ sequence but one inverted relative to the δ sequence termini of Ty1-B10. This rules out the necessity of a

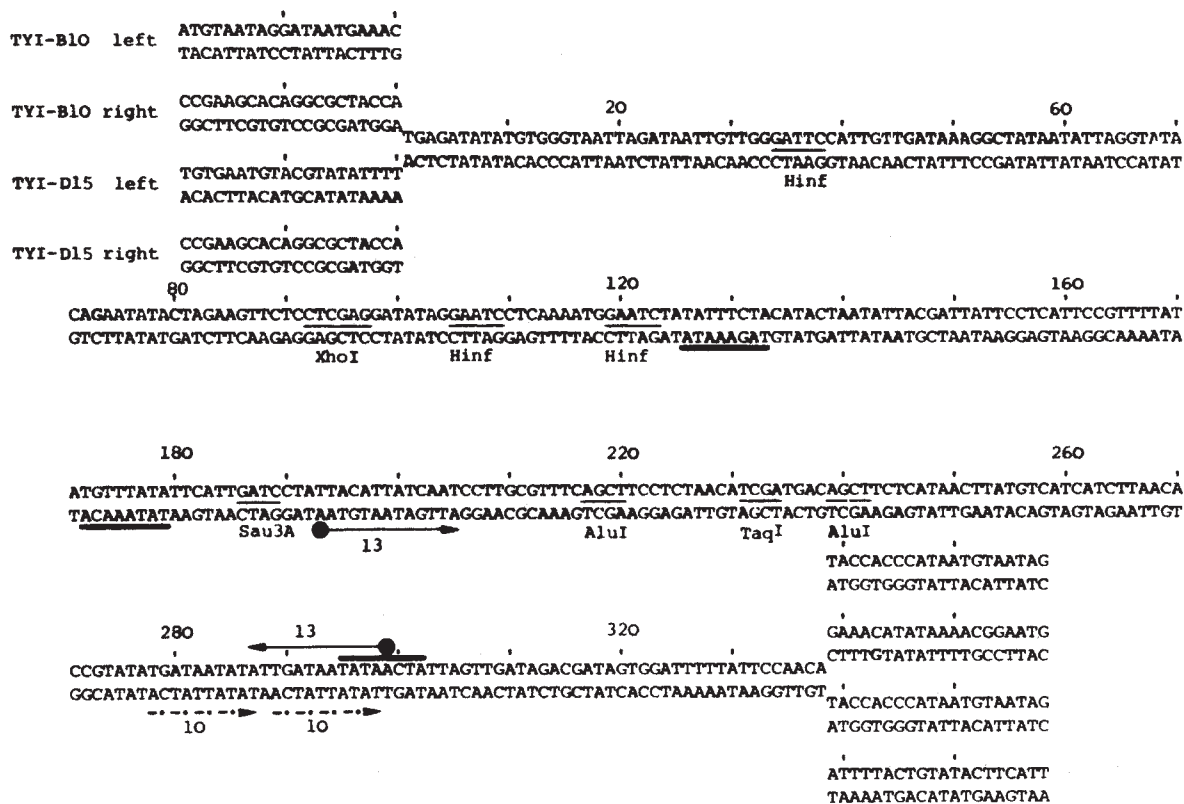


Fig. 3 Nucleotide sequences at the left and right termini of Ty1-B10 and Ty1-D15. The δ sequence of the four termini is written only once. The left termini start with a stretch of 20 nucleotides of the adjoining chromosomal DNA and continue at the end of the δ sequence with 20 nucleotides of ϵ DNA. The right termini start with 20 nucleotides of ϵ DNA and continue at the δ sequence end with 20 nucleotides of the adjoining chromosomal DNA. The three octanucleotides in the δ sequence which are very similar to the 'Hogness box' (5'-TATAAATA), common to eukaryotic promoter regions²⁸, are marked with a thick bar. Five other similar octanucleotides start at nucleotides 67 and 199 in the upper strand and at nucleotides 70, 143 and 171 in the lower strand. The longest open reading frame in the δ sequence consists of 42 codons (nucleotides 171-296 in the upper strand). One tandem repeat of 10 nucleotides and one inverted repeat of 13 nucleotides are indicated. The terminal sequence 5'-CCAACA appears inverted close to the beginning of the δ sequence (nucleotides 33-28). The adenosine phosphate at position 188 was difficult to read because of a very weak A signal at this position. The sequence of the opposite strand clearly showed a T at position 188. This was confirmed by cleavage with *Sau3A* restriction endonuclease. A poly-adenylation sequence is present at nucleotides 332-327.

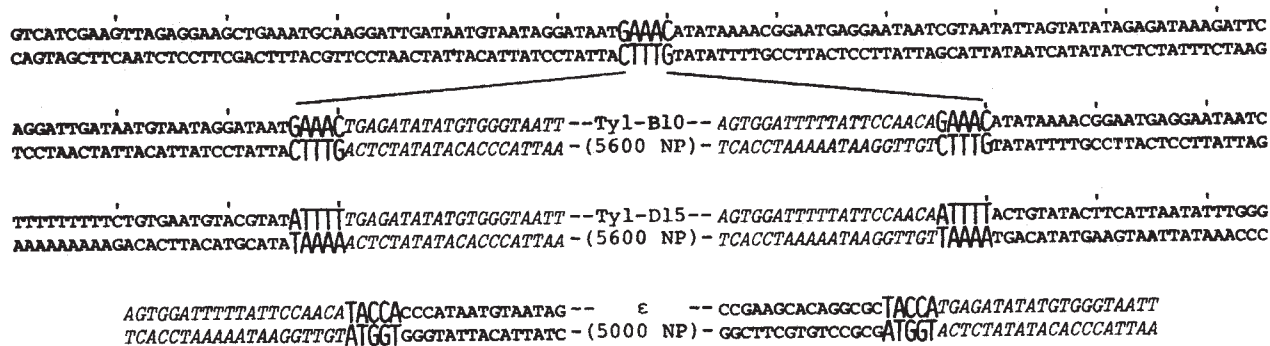


Fig. 4 Five nucleotide pairs duplication at δ sequence junctions. The first and second DNA sequences represent the same region left of the *Sup4* locus in strain S288C (no Ty1 inserted) and B596 (Ty1-B10 inserted), respectively. The relationship between the S288C and the B596 strain is unknown. The S288C sequence was obtained from both strands of a 130 nucleotide *TaqI* sequence. The third sequence shows the junction between the δ termini of Ty1-D15 and chromosomal DNA. The fourth sequence describes the internal Ty1 junctions between δ and ϵ DNA.

sequence homology between the Ty1 termini and the target DNA for Ty1 transposition. This is consistent with most of the sequence data from the insertion sites of bacterial transposons, although sequence homology might be involved in Tn3 transposition in the site selection but not the insertion process³².

It is unclear whether the 5-nucleotide deletion in the S288C sequence as compared with the inverted δ sequence of Fig. 3 is of any importance for the transposition. This deletion creates the 5'-GAAAC, the actual site of integration for Ty1-B10 (Fig. 4). The sequence without the deletion reads 5'-GAATA-

TAAAC (nucleotides 182-173 of Fig. 3). Possibly, this sequence alteration made the A+T-rich region in the solo δ sequence more receptive for a Ty1 transposition.

Our sequence data allow a comparison between an eukaryotic transposable element and the bacterial transposons on a molecular level. Shared features are the terminal repetition within the element, the duplication of target sequences at the sites of insertion and the lack of homologous recombination in the transposition process. We also noted some dissimilarities; the δ sequences at the Ty1 termini lack inverted repeats at their ends

and the 5-nucleotide internal duplication at the δ - ϵ juncture in Ty1 is not found at comparable positions in bacterial transposons. When compared with higher eukaryotes, Ty1 also shows certain similarities with the proviral DNA of retroviruses, such as the direct repetition of probably identical sequences of several hundred nucleotides at the two ends and the insertion into different chromosomal sites. As we learn more about eukaryotic transposable elements we may find universal features for DNA transposition processes in nature, although the various elements will surely have individual characteristics.

Received 14 March; accepted 16 May 1980.

- McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **21**, 197-216 (1956).
- Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (eds) *DNA Insertion Elements, Plasmids and Episomes* (Cold Spring Harbor Laboratory, New York, 1977).
- Kleckner, N. *Cell* **11**, 11-23 (1977).
- Starlinger, P. *Plasmid* (in the press).
- Calos, M., Johnsrud, L. & Miller, J. H. *Cell* **13**, 411-418 (1978).
- Grindley, N. D. F. *Cell* **13**, 419-426 (1978).
- Kühn, S., Fritz, H. J. & Starlinger, P. *Molec. gen. Genet.* **167**, 235-241 (1979).
- Johnsrud, L., Calos, M. P. & Miller, J. H. *Cell* **15**, 1209-1219 (1978).
- Oka, A., Nomura, N., Sugimoto, K., Sugisaki, H. & Takanami, M. *Nature* **276**, 845-847 (1978).
- Kleckner, N. *Cell* **16**, 711-720 (1979).
- Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. *Nature* **272**, 414 (1978).
- Ghosal, D., Sommer, H. & Saedler, H. *Nucleic Acids Res.* **6**, 1111-1122 (1979).
- Ohtsubo, H., Ohmori, H. & Ohtsubo, E. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1269-1278 (1979).
- Kamp, D. & Kahmann, R. *Nature* **280**, 247-250 (1979).
- Habermann, P., Klaer, R., Kühn, S. & Starlinger, P. *Molec. gen. Genet.* **175**, 369-373 (1979).

We thank Ron Davis, John Cameron, Elwyn Loh and Peter Shalit for helpful discussions and for recombinant λ phage, Werner Arber, Tom Bickle and Tom Fox for their comments on this manuscript. The work was supported by grant no. 3.330.78 from the Swiss National Foundation for Scientific Research.

Note added in proof: Sequence analysis of the *Drosophila* mobile element copia show that the integration of copia also generates a 5-nucleotide paired duplication and that the first three and the last ten nucleotides of the terminal repeats in copia and Ty1 are identical (B. Levis, P. Dunsmuir, W. Brorein & G. Rubin, personal communication).

- Gill, R. E., Heffron, F. & Falkow, S. *Nature* **282**, 797-801 (1979).
- Chou, J., Lemaux, P. G., Casadaban, M. J. & Cohen, S. N. *Nature* **282**, 801-806 (1979).
- Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16**, 739-751 (1979).
- Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. *Cold Spring Harb. Symp. quant. Biol.* **42**, 1053-1063 (1978).
- Tchurikov, N. A., Ilyin, Yu. V., Ananiev, E. V. & Georgiev, G. P. *Nucleic Acids Res.* **5**, 2169-2187 (1978).
- Strobel, E., Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 429-439 (1979).
- Potter, S. S., Brorein, W. J. Jr, Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 415-427 (1979).
- Gehring, W. & Paro, R. *Cell* (in the press).
- Hughes, S. H. et al. *Cell* **15**, 1397-1410 (1978).
- Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933-1937 (1979).
- Arthur, A. & Sherratt, D. *Molec. gen. Genet.* **175**, 267-274 (1979).
- Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
- Gannon, F. et al. *Nature* **278**, 428-433 (1979).
- Meyer, J., Iida, S. & Arber, W. *Molec. gen. Genet.* (in the press).
- MacHattie, L. A. & Jackowski, J. B. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 219-228 (Cold Spring Harbor Laboratory, New York, 1977).
- Iida, S. & Arber, W. *Molec. gen. Genet.* **177**, 261-270 (1980).
- Tu, C. P. D. & Cohen, S. N. *Cell* **19**, 151-160 (1980).
- Taylor, J. M. *Curr. Top. Microbiol. Immun.* **87**, 23-41 (1979).

Transcriptional control of two gene subclusters in the tRNA operon of bacteriophage T4

Alexander Goldfarb* & Violet Daniel

Biochemistry Department, Weizmann Institute of Science, Rehovot, Israel

In bacteriophage T4 DNA, transcription units recognized *in vitro* by host RNA polymerase consist of promoter-proximal 'immediate early' (IE) genes and promoter-distal 'delayed early' (DE) genes separated from each other by ρ -dependent transcription terminators¹. *In vivo*, the transition from IE to DE transcription requires phage-specific protein synthesis and can be prevented by chloramphenicol (CAM)². Most of the information about IE/DE transition has been obtained by hybridization analyses of mixtures of RNA species synthesized simultaneously on several T4 transcription units (for review see ref. 3). A useful model for the study of T4 gene expression at the level of primary transcripts and individual gene products is provided by the T4 tRNA operon, a cluster of genes coding for eight T4-specific transfer RNAs and two stable RNAs (species 1 and 2) of unknown function (Fig. 1). The 10 genes of the tRNA operon are arranged in two subclusters (I and II) with a promoter located about 1 kilobase pair upstream^{4,5}. The primary transcripts and the final gene products of this region have been identified and isolated^{6,7}. Moreover, this genetic region was recently cloned and a part of it sequenced^{5,8}. We describe here the expression of T4 tRNA genes *in vivo* and *in vitro* in terms of the IE/DE concept and demonstrate that the two subclusters of the tRNA operon are subject to different modes of control.

The transcription of individual tRNA genes *in vitro* can be monitored by digesting the total synthesized RNA with nucleases present in the 100,000g supernatant (S-100) of *Escherichia coli* and analysing the stable products by gel elec-

trophoresis^{7,9}. Figure 2 shows the S-100 digestion products of RNA synthesized in the presence and in the absence of ρ -factor. As templates for transcription in this experiment we used wild-type T4 DNA (lanes 1, 2) or DNA with deletions in the tRNA operon (lanes 3-6), the map locations of which are shown in Fig. 1. The digestion products migrating in the 4S region of lanes 1 and 2 (Fig. 2a) were analysed by a second-dimension electrophoresis (Fig. 2b). One can see that factor ρ prevents the transcription of genes coding for tRNA^{Arg} and RNA species 1 and 2 from the wild-type DNA. However, in the presence of ρ , these genes are expressed from DNA containing deletion Δ 119, and species 1 RNA is synthesized in the case of deletion Δ 27. The addition of ρ does not block the expression of genes of subcluster I. These results can only be explained by assuming the existence of a ρ -dependent transcription terminator in DNA between subclusters I and II of the tRNA operon. The absence of ρ -effect in the case of deletions Δ 119 and Δ 27 seems to be due to the removal of this site from DNA by the deletions.

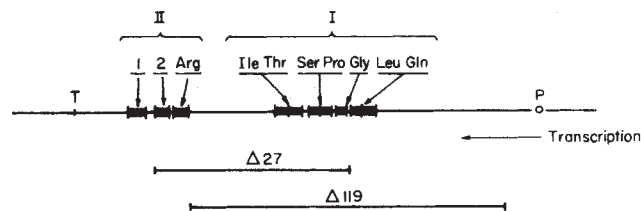


Fig. 1 T4 tRNA operon (modified from refs 4, 5, 7, 13). I and II, the two subclusters of stable RNA genes. The distance between promoter (P) and terminator (T) is about 3,000 base pairs. Lower bars represent deletions used in the present work.

The effect of ρ on the transcription of tRNA genes was further analysed at the level of primary transcripts. Elsewhere⁷ we have studied the *in vitro* transcription of this region at high ionic strength when ρ -factor is not active and demonstrated that the major transcript of the wild-type tRNA operon is a 3.0-kilobase long RNA molecule (RNA species A_{wt}) which contains all 10

* Present address: Max-Planck-Institute for Biochemistry, Martinsreid bei Munchen, FRG.

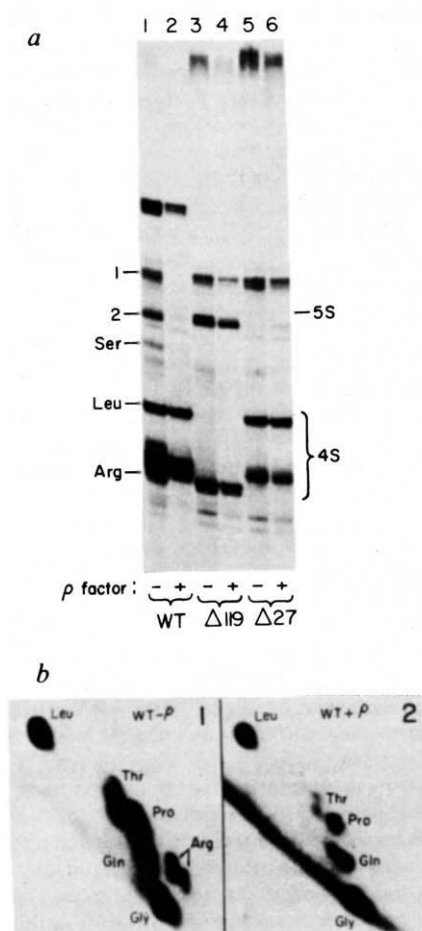


Fig. 2 Autoradiograms showing stable RNAs synthesized *in vitro* in the presence and absence of ρ -factor. DNA (10 μ g) from wild-type (WT) T4 or psu_{b} deletion mutants $\Delta 119$ and $\Delta 27$ were transcribed by 4 μ g RNA polymerase in 100- μ l reaction mixtures containing 40 mM Tris-HCl pH 7.9, 10 mM MgCl_2 , 7 mM 2-mercaptoethanol, 50 mM KCl, 0.25 mM each of ATP, GTP and UTP and 0.08 mM [α - 32 P]CTP (1 Ci mmol $^{-1}$, Amersham); 4 μ g of ρ -factor were added where indicated. After 15 min incubation at 37°C, 5 μ g ml $^{-1}$ rifampicin was added and incubation continued for 10 min followed by addition of 10 μ g DNase 1 (RNase-free, Worthington). After 10 min incubation the samples were deproteinized with phenol, RNA precipitated with ethanol, dried, dissolved and digested with 5 mg ml $^{-1}$ 100,000g supernatant fraction of *E. coli* (S-100) for 2 h at 37°C. The digestion products, after phenol deproteinization and ethanol precipitation, were separated by electrophoresis in acrylamide gels containing 6.6 M urea. The isolation of RNA polymerase and DNA, preparation of S-100, conditions of S-100 digestion and gel electrophoresis and identification of digestion products were as described elsewhere⁷. Factor ρ was purified according to ref. 14. *a*, Separation of S-100 digestion products in 10% acrylamide gel. *b*, Two-dimensional separation of 4S wild-type products. The 4S regions of lanes 1 and 2 of gel *a* were cut out and applied to 20% acrylamide gel slabs for the second-dimension electrophoresis.

genes of the operon and is terminated about 0.3 kilobases downstream from subcluster II. It was also shown that deletions $\Delta 27$ and $\Delta 119$ result in the synthesis of new deletion-specific transcripts A_{27} and A_{119} which are initiated and terminated at the same sites as the wild-type transcript A_{wt} (ref. 7). We now find (Fig. 3) that at low salt concentrations, transcripts A_{wt} , A_{27} and A_{119} are still synthesized on their respective DNA templates. The addition of ρ -factor results in the disappearance of the wild-type transcript A_{wt} whereas the two deletion-specific transcripts are still synthesized. This confirms the above conclusion about ρ -dependent termination of transcription between the two subclusters of tRNA operon.

On the basis of these observations one can expect that the genes of subcluster II would not be transcribed *in vivo* in the absence of phage-specific protein synthesis. Accordingly, we analysed the formation of phage-specific stable RNA in cells infected in the presence of CAM (Fig. 4). In this experiment we compared RNAs induced by wild-type phage (lanes 2, 4) and deletion mutant $\text{psu}_{\text{b}} \Delta 119$ (lanes 3, 5) in the presence (lanes 4, 5) and absence (lanes 2, 3) of CAM. Note that in the presence of CAM host transcription is not completely shut off by bacteriophage and the continued synthesis of bacterial tRNA gives a high background. Nevertheless, the products of subcluster II, tRNA^{Arg} and species 1 and 2, can be detected. It is clear that CAM strongly inhibits the formation of these three products in the cells infected with wild-type phage. In the case of deletion $\Delta 119$, however, the formation of RNA species 1 and tRNA^{Arg} is CAM resistant. The CAM sensitivity of RNA species 2 can be attributed to its degradation in the absence of the product of bacteriophage gene *mb*: the product of this early gene has been shown to be required for the appearance of tRNA^{Pro}, tRNA^{Ser},

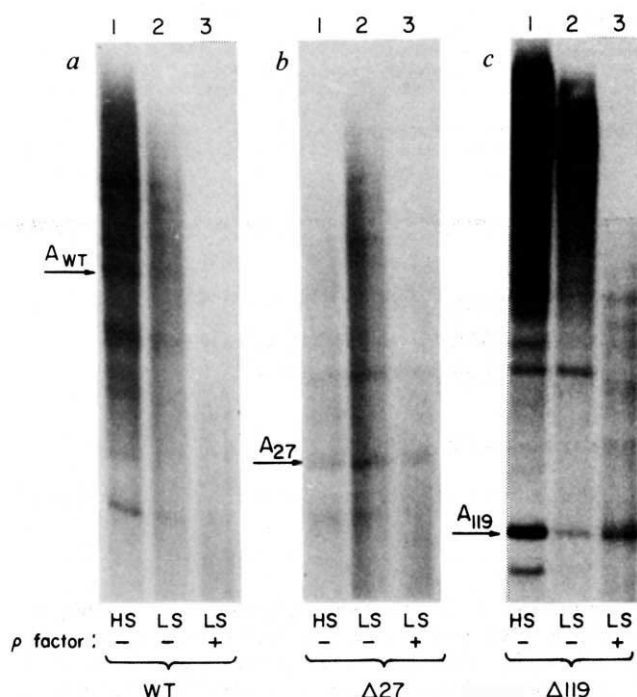


Fig. 3 Autoradiogram showing primary transcripts of T4 DNA made in the presence and absence of ρ -factor. DNA of wild-type T4 (WT) and deletion mutants $\text{psu}_{\text{b}} \Delta 27$ and $\Delta 119$ was transcribed at low (LS, 50 mM KCl) and high (HS, 250 mM KCl) salt concentrations. Otherwise, the conditions of transcription were as in Fig. 2. After the first phenol deproteinization, RNA transcripts were precipitated with ethanol, dissolved and separated in composite acrylamide-Agarose SDS gels as described elsewhere⁷. The lengths of RNA species A_{wt} , A_{27} and A_{119} are about 3,000, 1,500 and 750 nucleotides, respectively. *a*, *b*, 2.2% acrylamide-0.5% Agarose gels. *c*, 2.5% acrylamide-0.5% Agarose gel.

tRNA^{Ile} and RNA species 2 in the infected cell^{10,11}. Although, in the absence of *mb* function these RNAs are transcribed and probably processed, they cannot be detected *in vivo*, apparently due to their fast degradation. As for tRNA^{Arg} and RNA species 1, we conclude that phage-specific protein synthesis is required for their transcription from the wild-type template but not from DNA with deletion $\Delta 119$. This, as well as the above *in vitro* experiments, demonstrate that the genes of subcluster II are expressed as DE functions in the wild-type phage and as IE functions in $\text{psu}_{\text{b}} \Delta 119$ (and $\Delta 27$) mutant.

The CAM inhibition of subcluster II transcription can hardly be explained by the polar effect of nonspecific inhibition of translation³ because tRNA genes are not likely to be translated at all. Rather, one can assume that a phage-specific protein whose target sequence is situated between the two subclusters of tRNA genes is responsible for the transcription of subcluster II.

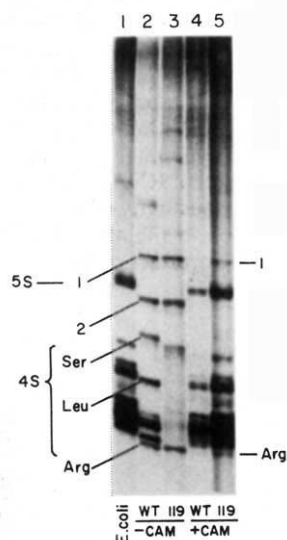


Fig. 4 Autoradiogram showing T4-specific stable RNAs induced in *E. coli* in the presence and absence of CAM. The cells of *E. coli* were infected with wild-type T4 (WT) or deletion mutant $\text{psu}_{\Delta 119}$ at MOI = 10 and labelled with ^{32}P as described elsewhere⁷. Where indicated, $100 \mu\text{g ml}^{-1}$ CAM was added to the culture 1 min before phage infection. RNA was extracted and separated in 10% acrylamide urea gel. Lane 1 shows stable *E. coli* RNAs.

The existence of a phage-specific product responsible for IE/DE transition has been postulated in T4 (ref. 3). Its action may be similar to the anti-termination effect of *N* gene product of bacteriophage λ (ref. 12). The study of such proteins has hitherto been hampered by the absence of suitable assay systems. Our results provide a convenient *in vitro* system for the search and analysis of the factors responsible for IE/DE transition.

We thank Drs K. Fukada and J. Abelson for $\text{psu}_{\Delta 119}$ mutants and for communicating their results before publication, and Mr Y. Tichauer for isolating ρ -factor. This research was supported in part by a grant from the US-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

Received 4 January; accepted 29 April 1980.

- O'Farrell, P. Z. & Gold, L. M. *J. biol. Chem.* **248**, 5512–5519 (1973).
- Salzer, W., Bolle, A. & Epstein, R. *J. molec. Biol.* **49**, 271–296 (1970).
- Rabussay, D. & Geiduschek, E. P. in *Comprehensive Virology* Vol. 8 (eds Fraenkel-Conrat, H. & Wagner, R.) (Plenum, New York, 1976).
- Fukada, K., Ostuka, A. & Abelson, J. *J. molec. Biol.* **137**, 191–211 (1980).
- Fukada, K., Gossens, L. & Abelson, J. *J. molec. Biol.* **137**, 213–234 (1980).
- Goldfarb, A., Seaman, E. & Daniel, V. *Nature* **273**, 562–564 (1978).
- Goldfarb, A. & Daniel, V. in preparation.
- Fukada, K. & Abelson, J. *J. molec. Biol.* (in the press).
- Nierlich, D. P., Lamfrom, H., Sarabhai, A. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 179–182 (1973).
- Wilson, J. H. & Abelson, J. *J. molec. Biol.* **69**, 57–73 (1972).
- McClain, W. H., Guthrie, C. & Barrell, B. G. *J. molec. Biol.* **81**, 157–171 (1973).
- Adhya, A. & Gottesman, M. A. *Rev. Biochem.* **47**, 967–996 (1978).
- Wilson, J. M., Kim, J. S. & Abelson, J. *J. molec. Biol.* **71**, 547–556 (1972).
- Roberts, J. W. *Nature* **244**, 1168–1174 (1969).

Evolution of DNA sequences has been retarded in Malagasy primates

Tom I. Bonner*, Robert Heinemann† & George J. Todaro*

*Laboratory of Viral Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

†Program on Genetics, George Washington University, Washington, DC 20052

It is generally accepted¹ that there are six major groups of living primates: (1) lemurs (including all the primates of Madagascar), (2) lorises (including galago and potto), (3) tarsiers, (4) New World monkeys, (5) Old World monkeys and (6) apes (including man). Tree shrews, once considered to be primates, are now generally recognized² as not significantly more closely related to the six groups than other mammals. The first surviving primate lines to diverge from the common primate ancestor are believed to have given rise to one or more of the first three groups. However, the fossil record is insufficient to determine their relative branching order³. Furthermore, neither morphological considerations^{2,4–7} nor studies of protein evolution^{8,9} produce unanimity as to whether tarsiers are more closely related to the prosimians (the lemurs plus lorises) or the simians (the monkeys and apes). In an attempt to resolve these discrepancies, we have measured the DNA sequence difference between several primates. We report here that the evolution of DNA of primates from Madagascar is significantly less than that of all other groups of living primates. This is not expected in the simplest form of the theory of neutral selection and may be important for our understanding of evolution at the molecular level.

To measure genetic differences between the various primates as directly as possible, we followed Kohne *et al.*¹⁰ in using the sequences which are present in a frequency of one copy per haploid genome, commonly referred to as single copy DNA. As the single copy fraction amounts to approximately half the DNA in the genome, the evolutionary fluctuations which can occur within a single gene are thoroughly averaged out. To determine the branching order in the early evolutionary history of the primates, we divide the living primates into four groups: lemurs, lorises, tarsiers and simians—New World plus Old World monkeys and apes. Our experiment consists of preparing ^{125}I -labelled single copy DNA from a representative species of each of the four groups, hybridizing each labelled DNA to the unlabelled total DNA of two to four members of each group in conditions such that each labelled DNA molecule hybridizes to an unlabelled molecule, and measuring the sequence divergence between the two species of DNA by measuring the melting temperature of the hybrid molecules. From experiments with various model systems, we can expect a 1 °C decrease in melting temperature for each ~1% nucleotide difference¹¹. Thus, what we measure is the accumulation of single base mutations in the two lineages since their divergence. We expect that these accumulated mutations account for the bulk of the evolutionary difference between the species. We will not, however, measure any additions to, deletions from or rearrangements of the DNA. Thus, we may not detect some of the changes in gene regulation which would be expected to be relatively rare but evolutionarily quite significant^{12,13}.

Our results are shown in the form of melting curves (Fig. 1). From these it is clear that with labelled DNA from the simian, loris or tarsier groups, the most closely related DNAs are always from the lemur group. The only explanations for this are that either the data are defective or the lemurs have diverged substantially more slowly than the other groups. As the melting curves for all the DNAs within a single group are nearly identical when hybridized with ^{125}I -DNA from another group, the only

possible defect would be one which simultaneously affects all the DNAs of a single group. This would require either the degradation of all the DNAs within one group or the presence of an impurity in some of the groups of DNAs that affects the stability of hybrids with some but not all of the other groups. The fact that all the homologous melting curves are nearly identical demonstrates that the DNAs of all four groups are comparable. When, as a further test, separate extractions (obtained from B. Hoyer) of DNAs from *Lemur catta* and tarsier were compared with those used for Fig. 1, identical results were obtained. Thus, we believe that the data can only be accounted for by the hypothesis that the lemur DNA has evolved more slowly than the other primate DNAs.

To construct a phylogenetic tree consistent with the data we must consider Fig. 1 quantitatively. If we define the hybridizable fraction of a given labelled DNA as the fraction which hybridizes to the homologous DNA, the data of Fig. 1 show that only 50–70% of the hybridizable DNA actually hybridizes to DNAs from other groups. The most reasonable explanation of this reduced hybridization is that a portion of the hybridizable sequences have diverged so much between the two groups that they are not able to form stable hybrids at the hybridization temperature. We can take these sequences into account if we use the melting temperature of the potentially hybridizable sequences rather than those which actually hybridize¹⁰. We therefore characterize each melting curve with the temperature (T_{50H}) at which 50% of the hybridizable sequences are bound to hydroxyapatite. These values are presented in Table 1. We then convert the measured T_{50H} values to the actual number of nucleotide substitutions which have occurred since the divergence of the two species. There is some uncertainty in this conversion as results from different model systems vary over

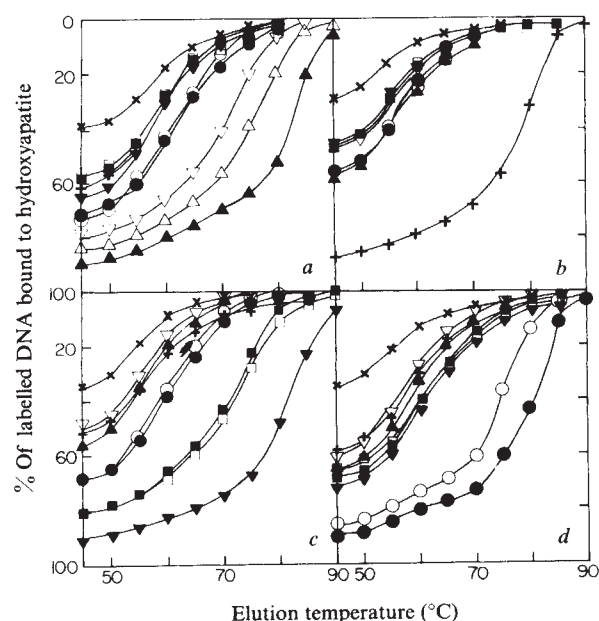


Fig. 1 Melting curves of hybrid DNA molecules formed between ^{125}I -labelled single copy DNAs and unlabelled cellular DNAs. DNAs were iodinated to a specific activity of 3×10^7 c.p.m. per μg by a modification of the procedure of Chan *et al.*²⁰, hybridized to a C_0t of 200 and fractionated on hydroxyapatite at 45°C in 0.14 M phosphate buffer (PB, sodium phosphate, pH 6.8). The single copy fraction, that is, the fraction which did not bind, was then fractionated by electrophoresis on an alkaline 2% agarose gel and the portion between 200 and 1,200 nucleotides was used for the hybridizations. 100 μg of cellular DNA was hybridized with about 2,000 c.p.m. of labelled DNA in 1.0 M PB at 60°C to a C_0t of 22,000, the hybrid was bound to hydroxyapatite in 0.14 M PB plus 0.02% SLS at 45°C and eluted in 5°C steps as described elsewhere²¹. The cellular DNAs are: $\blacktriangle$, human; $\triangle$, baboon; ∇ , spider monkey; $+$, tarsier; $\blacksquare$, galago; $\square$, slow loris; $\bullet$, slender loris; $\circ$, *L. catta*; $\odot$, *L. mongoz*; $\times$, pig or mouse. The labelled DNAs are: a, human; b, tarsier; c, galago; d, *L. mongoz*.

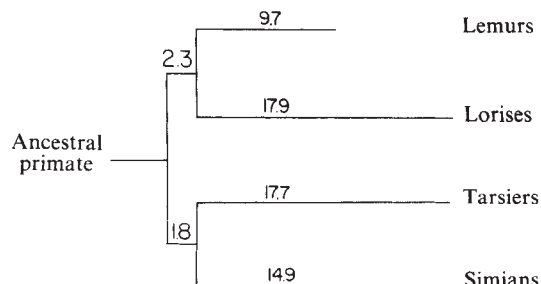


Fig. 2 Phylogenetic tree for the primates derived from the single copy DNA sequence divergences. The number of actual nucleotide substitutions since the divergence of the lineages as obtained from Table 2 were used to construct the tree as follows. The sum of the lemur and loris branches is given by the average of their reciprocal hybridizations. The difference between the lemur and loris branch lengths is obtained from the experiments in which labelled tarsier or simian DNA is hybridized to both lemur and loris DNAs (the differences within columns of Table 2). These values would be expected to be more accurate than the differences obtained by hybridizing labelled lemur and labelled loris DNAs to either simian or tarsier DNAs (the differences within rows of Table 2). Combining these values, we can solve for the length of the two branches. Similarly, we can solve for the length of the tarsier and simian branches. Using these four branch lengths we can then determine the length of the branch interconnecting the lemur-loris branch point and the tarsier-simian branch point necessary to fit the eight measurements which include this branch. Finally, these branch lengths are combined with the estimated distances measured with tree shrew DNA to determine the branch point with the common primate ancestor. The numbers on the branches represent the percentage of nucleotide substitutions along that branch.

$1.5\text{--}0.7^\circ\text{C}$ in melting temperature reduction per per cent nucleotide difference. We will assume the intermediate value of 1.0°C in our calculations. However, choice of any conversion factor within the above range does not significantly affect the relative amounts of change along the different branches. An additional correction must be made because the degree of sequence divergence in our measurements is sufficiently great for the probability of multiple changes at the same site to be appreciable. Assuming only that a given base can mutate to any other base with equal probability, statistical corrections¹⁴ can be made (see Table 2) and, as we are comparing about 10^9 nucleotides, the corrected values should reflect the actual number of base substitutions quite accurately. Once this correction is

Table 1 Sequence divergence as measured by reduction of hybrid melting temperature

Unlabelled DNA	Labelled single copy DNA			
	Human	Tarsier	Galago	<i>L. mongoz</i>
Human	0	24.6	29.3	23.4
Baboon	6.9	—	27.8	25.1
Spider monkey	11.2	26.4	30.7	24.5
Tarsier	26.2	0	30.2	26.9
Galago	26.7	30.1	0	22.1
Slow loris	27.6	—	11.5	23.0
Slender loris	28.6	28.6	11.3	23.1
<i>L. catta</i>	22.6	24.6	22.4	6.0
<i>L. mongoz</i>	21.2	24.6	22.0	0

The T_{50H} for each melting curve in Fig. 1 and for other melting curves which are not shown was determined as described in the text. The reduction in T_{50H} compared with the homologous DNA was averaged and this average is tabulated. The homologous T_{50H} values are 78.8–81.3, 78.1, 78.8–80.7 and 79.5–81.1 $^\circ\text{C}$ for human, tarsier, galago and lemur DNAs, respectively. The values for comparable pairs of DNAs are very close to those reported by Kohne, Chiscon and Hoyer¹⁰. The reciprocal values (for example, the divergence between human and tarsier as measured with either labelled tarsier or labelled human DNAs) are in general but not precise agreement. This difference probably results from the various labelled single copy DNAs not being precisely equivalent in sequence composition.

Table 2 Difference matrix in terms of per cent actual nucleotide substitutions

Unlabelled DNAs averaged within groups	Labelled single copy DNA			
	Simian	Tarsier	Loris	Lemur
Simian	0	32.0	38.2	30.1
Tarsier	33.0	0	39.7	34.3
Loris	35.2	38.2	0	28.1
Lemur	26.6	30.6	27.1	0

The values of ΔT_{50H} from Table 1 were averaged over the various unlabelled DNAs within a taxonomic group and were converted to per cent actual nucleotide substitutions. The conversion assumes that 1 °C in ΔT_{50H} equals 1% sequence difference and that any given base can mutate to any other base with equal probability. The use of Holmquist's¹⁴ equation (20) then corrects for multiple changes at a particular site in one lineage and for changes at the same site in both lineages.

made, it is clear that the total number of base substitutions along individual lineages should be additive. We can then use the six pairwise differences between the four primate groups to determine distances along each branch of the tree.

The only tree consistent with the data but without substantially negative branch lengths places the loris and lemur groups at one branch point separated by about 4% base substitution from the branch point between tarsiers and simians. It is not, however, possible to determine the location of the branch point for the divergence from the common primate ancestor without data which include a more distant DNA. We have therefore used single copy tree shrew (*Tupaia glis*) DNA which is so distant that we cannot directly measure ΔT_{50H} . We can, however, measure ΔT_{25H} values of 24.8, 26.8 and 28.0 °C for lemur, simian and loris groups, respectively. As Fig. 1 indicates that $\Delta T_{50H} = 1.18 \times \Delta T_{25H}$, we can extrapolate to obtain approximate ΔT_{50H} values of 29.3, 31.6 and 33.0 °C. When combined with the branch lengths derived in Fig. 2, these values indicate that the common ancestor branch point is located between the lemur-loris branch point and the simian-tarsier branch point. This position is consistent with both protein phylogenetic trees^{8,9}. The full branching diagram can then be drawn as shown in Fig. 2. With the exception of the length of the lemur and simian branches, the tree is qualitatively quite similar to that of Dene *et al.*⁸. In particular, the first branching is that of the lemur-loris ancestor from the tarsier-simian ancestor, followed shortly after by further splitting in both these branches to give rise to the four separate groups. The slight shortening of the simian branch relative to the loris and tarsier branches may reflect the fact that simian generation times are two to four times greater than those of prosimians. Whether the extent of molecular evolution depends strictly on time or on the number of generations remains unclear^{10,15}. The present data seem to support the latter hypothesis. However, the observation that lorises and tarsiers are approximately equidistant from the common primate ancestor whereas lemurs are nearly twice as close must have a different explanation because all three groups have comparable generation times. We believe that this result constitutes the first case in which a phylogenetic tree constructed using single copy DNA has branches of substantially unequal lengths.

As the *Lemur mongoz* and *L. catta* DNAs represent only one genus within the broader lemur group, the question arises as to how widespread the reduced divergence is within the lemur group as a whole. We have therefore tested the DNA from three additional genera of Malagasy primates, *Microcebus*, *Cheirogaleus* and *Propithecus*. Except for *Daubentonina*, the protein phylogenies^{8,9} place *Microcebus* and *Cheirogaleus* as the most distant Malagasy relatives of the genus *Lemur* and place *Propithecus* approximately half as distant from *Lemur*. Hybridization with various labelled single copy DNAs indicates that the *Propithecus* branch length is the same as for the genus *Lemur* and that the *Microcebus* and *Cheirogaleus* branches are only slightly longer (that is, 12.5% actual substitution). Thus,

reduced divergence rates occur throughout the lemur group. Similar evidence can be derived from the protein data of Sarich and Cronin⁹. Of 10 genera of Malagasy primates tested, 7 have branch lengths which are 61–83% of the loris group branch length. Considering all the evidence it is clear that essentially all the primates from Madagascar have had a reduced evolutionary rate since their divergence from the lorises.

Our results demonstrate that lemurs are genetically the most primitive of the modern primates in the sense that their DNAs most closely resemble that of the common primate ancestor. Although lemurs are considered quite primitive morphologically¹, this is unexpected on a molecular basis. Kimura¹⁶ has argued that the genetic load imposed by the observed rate of DNA divergence is so great that only a few per cent of the DNA could be subject to natural selection and that most mutations must therefore be of neutral selective value. In this case the rate of fixation of base substitutions should be proportional to the mutation rate¹⁶ and all the branch lengths would be expected to be equal, whereas in fact the average rate of DNA divergence for lemurs during the 50–70 Myr since their divergence from the lorises is approximately half that of the other prosimians. The reason for this discrepancy may be either that the bulk of the DNA is subject to selection or that the mutation rate is reduced for lemurs. The fact that this effect occurs for all Malagasy primates and not for their closest relatives, the lorises, implies that it is either an environmental effect or the result of a major genetic change which occurred after the divergence of the lemurs from the lorises and before the divergence of the various lemurs from each other.

There are three possible solutions. First, the basic premise of the theory of neutral selection could be wrong and most of the DNA could be subject to selection at the nucleotide level. If this is true, it follows that the Malagasy environment has required less adaptation than the environments experienced by other primates. Alternatively, the Malagasy primates could have a reduced mutation rate due to a change in the genes which affect the spontaneous mutation rate. This could happen in a somewhat trivial but strictly neutralist fashion as the result of a founder effect. If the lemurs are descended from a small number of ancestors which populated Madagascar^{17–19}, these individuals could have fortuitously represented a rare low mutation rate variant which was then perpetuated on Madagascar. In the absence of a founder effect, a reduced mutation rate in lemurs could also result from selection for an optimum spontaneous mutation rate without direct selection on the genome as a whole. As in the case of selection on the whole genome, this hypothesis would require that there has been a reduced selective pressure for change on Madagascar.

The selection and founder effect hypotheses could be evaluated by looking for the same effect in other orders of mammals present on Madagascar, such as carnivores or rodents. Assuming that such experiments do not establish the founder effect hypothesis, an understanding of the basis for the reduced evolutionary rate of the primates of Madagascar could resolve the question of whether darwinian selection or selective neutrality is the major force in molecular evolution.

We thank Dr Morris Goodman for suggesting these measurements, for his continuing help and for some of the prosimian tissues, especially the tarsier tissue which was made available to him by the Chicago Zoological Society. We also thank Dr Bill Hoyer for helpful discussions and for tarsier and *L. catta* DNAs, and the Duke University Center for the Study of Primate Biology and History for *Propithecus*, *Cheirogaleus* and *Microcebus* tissues. This work is in partial fulfilment of the requirements for the Ph.D degree to be submitted to the George Washington University by R.H.

Received 9 January; accepted 1 May 1980.

1. Charles-Dominique, P. & Martin, R. D. *Nature* **227**, 257–260 (1970).
2. Martin, R. D. in *Recent Advances in Primatology* Vol. 3 (eds Chivers, D. S. & Josey, K. A.) 3–26 (Academic, London, 1978).
3. Simons, E. L. in *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates* (eds Goodman, M. & Tashian, R. E.) 35–62 (Plenum, New York, 1976).

4. Luckett, W. P. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 157–182 (Plenum, New York, 1975).
5. Schwartz, J. H. in *Recent Advances in Primatology* Vol. 3 (eds Chivers, D. S. & Josey, K. A.) 195–202 (Academic, London, 1978).
6. Noback, C. R. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 199–218 (Plenum, New York, 1975).
7. Martin, R. D. *Symp. zool. Soc. Lond.* **33**, 301–337 (1973).
8. Dene, H. T., Goodman, M. & Pyrchodko, W. in *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates* (eds Goodman, M. & Tashian, R. E.) 171–195 (Plenum, New York, 1976).
9. Sarich, V. M. & Cronin, J. E. in *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates* (eds Goodman, M. & Tashian, R. E.) 141–170 (Plenum, New York, 1976).
10. Kohne, D. E., Chiscon, J. A. & Hoyer, B. H. *J. hum. Evol.* **1**, 627–644 (1972).
11. Bonner, T. I., Brenner, D. S., Neufeld, B. R. & Britten, R. J. *J. molec. Biol.* **81**, 123–135 (1973).
12. King, M.-C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
13. Britten, R. J. & Davidson, E. H. *Q. Rev. Biol.* **46**, 111–138 (1971).
14. Holmquist, R. J. *J. molec. Evol.* **1**, 115–133 (1972).
15. Sarich, V. M. & Cronin, J. E. *Nature* **269**, 354 (1977).
16. Kimura, M. *Nature* **217**, 624–626 (1968).
17. Walker, A. in *Calibration of Hominoid Evolution* (eds Bishop, W. W. & Miller, J. A.) 247–272 (Scottish Academic, Edinburgh, 1972).
18. Martin, R. D. *Phil. Trans. R. Soc. B264*, 295–352 (1972).
19. Tattersall, I. *Trans. N.Y. Acad. Sci.* **35**, 314–324 (1973).
20. Chan, H. C., Ruyechan, W. T. & Wetmur, J. G. *Biochemistry* **15**, 5487–5490 (1976).
21. Bonner, T. I. & Todaro, G. J. *Virology* **103**, 217–227 (1980).

Neutron scattering studies of *lac* repressor

M. Charlier*, J. C. Maurizot* & G. Zaccari†

* Centre de Biophysique Moléculaire, 1A, Avenue de la Recherche Scientifique, 45045 Orleans Cedex, France

† Institut Max von Laue-Paul Langevin, Avenue des Martyrs, BP, 156 Centre de Tri, 38042 Grenoble Cedex, France

The *lac* repressor, a tetrameric protein of identical subunits [molecular weight (MW) $4 \times 38,500$], interacts specifically with the *lac* operator, preventing the expression of the structural genes of the *lac* operon^{1,2}. The presence of an inducer (such as isopropyl- β -D-thiogalactoside; IPTG) which binds to the repressor, prevents operator binding by lowering the association constant by a factor of 10^3 (ref. 3). Genetic and biochemical analysis have shown that the major part—if not all—of the binding site for the *lac* operator is located in the 60 N-terminal residues of the protein^{4,5}. In certain conditions, limited trypsinolysis of the protein yields four N-terminal 'headpieces' each containing 51 or 59 residues, and a tetrameric core with full inducer binding activity^{4,5}. It was shown recently that this headpiece is able to bind nucleic acids, and interacts with the *lac* operator⁶, giving the same pattern of sensitivity with respect to the methylation of the bases as does the intact repressor⁷. We are studying the interaction of *lac* repressor with DNA by neutron scattering using contrast variation and discuss here measurements on the protein, its tryptic core and their complexes with IPTG. Our results demonstrate that the headpieces are located far (67 ± 10 Å) from the centre of mass of a somewhat elongated core, and that the inducer does not significantly change the radius of gyration of the protein.

Guinier plots⁸ of *lac* repressor and its tryptic core are shown in Fig. 1. Two parameters were obtained from the Guinier plots—a radius of gyration of excess scattering length density, R_G , and the scattered intensity extrapolated to zero angle $(I/C)_0$, which is related to the MW and concentration C of the particle⁹. In this case, the MWs of the particles are well known¹⁰ and the concentrations were measured by absorption spectrophotometry. These values provided a criterion for each measurement of $(I/C)_0$, as aggregation or interparticle effects would lead to misleading observations of both $(I/C)_0$ and R_G . Aggregation was not negligible above concentrations of about 5 mg ml^{-1} and appeared more readily in D_2O solvents, an observation which has also been made in other systems (B. Jacrot and G. Zaccari, personal communication). Values of R_G and $(I/C)_0$ measured in different conditions are shown in Table 1. The difference between R_G in H_2O and D_2O is expected

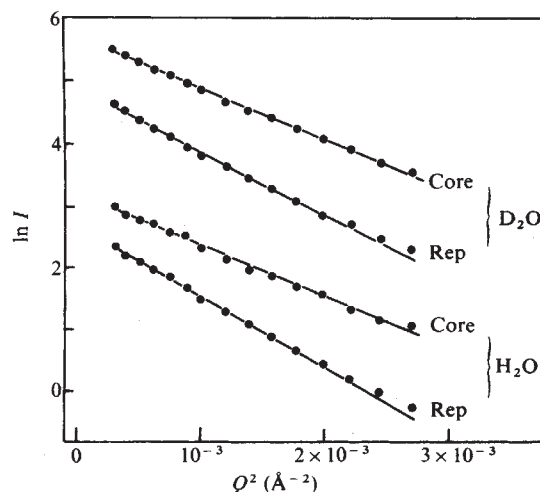


Fig. 1 Guinier plots of the scattering data for the *lac* repressor and its tryptic core. The logarithm of the scattered intensity is plotted against Q^2 ; $Q = (4\pi \sin \theta)/\lambda$, where θ is the half of the scattering angle and λ is the wavelength of the incident neutrons (10 Å in these measurements). Data were obtained with the protein concentrations 2.0 mg ml^{-1} (rep) and 2.5 mg ml^{-1} (core) in D_2O , and 5.0 mg ml^{-1} (rep) and 6.6 mg ml^{-1} (core) in H_2O . Buffers are described in Table 1 legend. The *lac* repressor was isolated by the method of Müller-Hill *et al.*²⁴. The tryptic core was prepared according to Files and Weber⁵. For clarity, the curves are shifted arbitrarily along the ordinate axis.

because of the hydrophilic amino acid residues lying predominantly on the protein surface⁹. There seems to be no change in R_G on binding of IPTG.

The radius of gyration of a compact protein such as lysozyme or haemoglobin, for example, is proportional, to a good approximation, to the cube root of its molecular weight¹¹. By extrapolating, we find the radii of gyration of compact multi-subunit proteins of MWs 128,000 and 154,000 to be about 30 Å and 32 Å, respectively. A higher value is observed for the tryptic core, and the experimental R_G of *lac* repressor is considerably larger. A larger R_G could indicate a less compact distribution of

Table 1 Neutron scattering data for *lac* repressor and its tryptic core

	Repressor	Core
MW	154,000	128,000
Volume (Å ³)	196,700	163,800
R_G HKPD (Å)	42.5 ± 1.5	35.0 ± 1.5
R_G DKPD (Å)	39.0 ± 0.2	34.2 ± 0.2
R_G DKPD + IPTG (Å)	39.2 ± 0.3	34.2 ± 0.2
$(I/C)_0$ obs. HKPD	0.0750 ± 0.0050	0.0620 ± 0.0010
$(I/C)_0$ calc.	0.0733	0.0608
D_{max} (Å)	115 ± 15	100 ± 15

MW and volumes are calculated from the sequence of the protein¹⁰ using data given in ref. 9. R_G and $(I/C)_0$ are, respectively, the radius of gyration and the scattered intensity at 0 angle per mg per ml of protein normalized by the incoherent scattering of H_2O . The values are obtained from Guinier plots. $(I/C)_0$ calculated (calc.) is obtained from the formula

$$(I/C)_0 = 4\pi 10^{-3} \times (T/1 - T) \times t \times N_A \times [\sum b/MW - \rho_s \times \bar{V} \times 10^{-24}/N_A]^2 \times MW$$

where T is the neutron transmission of the sample in H_2O buffer, t the path length in cm, N_A the Avogadro's number, $\sum b$ the sum, in cm, of the scattering length of the amino acid residues (given in ref. 9), $\bar{V}$ the partial specific volume, ρ_s the neutron scattering density of water in $\text{cm}^3 \text{ Å}^{-3}$ and MW the molecular weight¹². HKPD buffer is 0.2 M potassium phosphate and 10^{-4} M dithiothreitol, pH 7.25, in H_2O ; DKPD buffer is 0.2 M potassium phosphate, 0.5 M potassium chloride, 10^{-4} M dithiothreitol, pH 7.25 in D_2O . The concentration of inducer IPTG (10^{-3} M) is large enough to complex all repressor or core molecules ($K_a = 10^6 \text{ M}^{-1}$). D_{max} is the maximum dimension of the particle, calculated from Fourier transform of the scattering curves.

scattering mass within a similar shape, or a more elongated shape.

By examining the scattering curves at large angles, a 'scattering equivalent' to the shape of a particle can be deduced. It is a simple shape (an ellipsoid or a cylinder of given axial ratios, for example) whose calculated scattering curve best approximates the observed curve. Scattering curves of *lac* repressor and its tryptic core are shown in Fig. 2. Up to $R_G \times Q \approx 3$, the observed points fit very well to the curve calculated for a 2:1:1 prolate ellipsoid. The shape of compact proteins usually lies between that of a sphere and a 2:1:1 prolate ellipsoid. It is likely, therefore, that the large observed radius of gyration arises from a less compact packing of the subunits. The headpieces must be located so as not to change significantly the apparent shape of the particle. Although the data are poor in that region of Q , the shoulders in the scattering curves seem to be higher in D_2O buffers than in H_2O buffers. A similar observation has been discussed with respect to the packing of subunits in bacterial F_1 -ATPase¹². These shoulders must arise from a pronounced inhomogeneity in the scattering density distribution in the particle, probably associated with the packing of the subunits. The shoulders cannot be fitted by the scattering curves of homogeneous ellipsoids. Their heights as well as the angle at which they occur are approximately correct for hollow ellipsoids. Because the shoulder is higher in D_2O , we can say that the inhomogeneity is more pronounced in that solvent, as would be the case if there were solvent regions lined with hydrophilic amino acid residues 'within' the general ellipsoidal shape of the particle.

The maximum dimensions of the particles were deduced from the scattering curves in D_2O (the data from D_2O are better than those from H_2O solvents because of increased signal-to-background ratio) using the characteristic function given in ref. 8:

$$\gamma_0(r) = C' \times \int_0^\infty Q^2 I(Q) \frac{\sin Qr}{Qr} dQ$$

They are $115 \pm 15 \text{ \AA}$ for *lac* repressor and $100 \pm 15 \text{ \AA}$ for its tryptic core.

Assuming no major rearrangement on trypsin digestion, as supported by several experiments¹³⁻¹⁵, the large R_G of *lac* repressor shows that the headpieces extend outwards from the core particle. The distance d between the centre of mass of each

headpiece and the centre of mass of the protein can be calculated using the relationship (parallel axes theorem)

$$MR_G^2 = M'R_G'^2 + 4mr_G^2 + 4md^2$$

where M , M' and m are the masses of the repressor, the tryptic core and the headpiece, respectively, and R_G , R'_G and r_G are their respective radii of gyration. The calculation can be made for two extreme hypotheses concerning the conformation of the headpiece: either a spherical compact headpiece whose MW (6,400) leads to a radius of gyration of $10 \text{ \AA}$, or an elongated ellipsoid with a long axis of $35 \text{ \AA}$ so as to cover half the operator, for example¹⁶, leading to $r_G = 16 \text{ \AA}$. From these values, the calculated distances between the centre of mass of the protein and that of the headpieces are $67.5 \pm 10 \text{ \AA}$ and $66.5 \pm 10 \text{ \AA}$, respectively. This calculation seems to be very insensitive to the assumed value of the radius of gyration of the headpiece. Combining these results with data on the structure of the headpieces¹⁷ and its mobility^{18,19}, a model emerges with the headpieces as partly compact bodies at the end of flexible sequences attaching them to the core.

Two models have been proposed for the structure of *lac* repressor, based on packing considerations in microcrystals²⁰ and electron micrographs²¹. Our results do not allow us to confirm or reject either of these models. The size given by the authors is too uncertain to calculate valid radii of gyration and scattering curves (for example, the radii of gyration calculated from the extreme values given by Ohshima *et al.* are 38.4 and $45.5 \text{ \AA}$)²¹. The situation of the headpieces at about $68 \text{ \AA}$ from the centre of mass implies an outward localization of these parts of the protein. Because at least two subunits are necessary to bind the operator^{22,23} and because chemical cross-links between two headpieces have been observed²³, the headpieces must be arranged in pairs. In the elongated model proposed by Steitz *et al.*²⁰ the headpieces should be located in pairs near the extremities of the molecule. In this case, it is probable that the operator interacts with one pair of adjacent headpieces located at the same extremity of the repressor. In the model proposed by Ohshima *et al.*²¹ more possibilities of localization exist; however, in any case, the headpieces must be located on the outside of the proposed volume. In this latter case the headpieces probably do not appear clearly in the electron micrographs after negative staining. Repressor could bind to the operator with the two adjacent interacting headpieces fitting perfectly the sequence of bases and the conformational state of the DNA. When the inducer is present the relative positions of the headpieces might vary in such a way that the positions of the interacting residues no longer fit the operator sequence exactly, to result in a decrease of three orders of magnitude in the association constant with DNA. This movement of the headpieces on IPTG binding might be small as it must not change the radius of gyration significantly. On the other hand, it could be quite large provided the distance between the headpieces and the centre of mass of the particle did not change; such a displacement would not change the radius of gyration at all. Experiments are in progress to test this model.

We thank C. Hélène and B. Jacrot for encouragement and useful discussions, and W. Gilbert for communication of unpublished results. Part of this work was supported by the Délégation Générale à la Recherche Scientifique et Technique (contract 79-7-0150 to J.C.M., ACC Mécanisme de Reconnaissance Moléculaire).

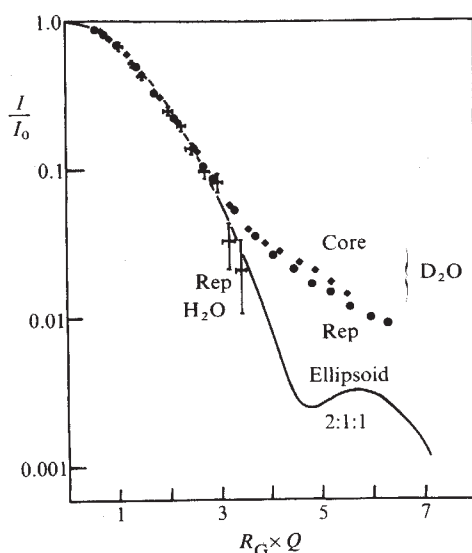


Fig. 2 Scattered intensity for large values of Q . Data are plotted against $R_G \times Q$, to compare directly the shape of normalized curves. Circles, diamonds and crosses are experimental values obtained for the indicated samples. The continuous line is the calculated curve for a prolate ellipsoid 2:1:1, according to ref. 8. Buffers are described in Table 1 legend. Protein concentrations are 5 mg ml^{-1} (rep) and 3.5 mg ml^{-1} (core) (neutron wavelengths $5 \text{ \AA}$).

Received 17 December 1979; accepted 14 May 1980.

1. Müller-Hill, B. *Prog. Biophys. molec. Biol.* **30**, 227-252 (1975).
2. Bourgeois, S. & Pfahl, M. *Adv. Protein Chem.* **30**, 1-99 (1976).
3. Barkley, M. D., Riggs, A. D., Jobe, A. & Bourgeois, S. *Biochemistry* **14**, 1700-1712 (1975).
4. Platt, T., Files, J. G. & Weber, K. *J. Biol. Chem.* **248**, 110-121 (1973).
5. Files, J. G. & Weber, K. *J. Biol. Chem.* **251**, 3386-3391 (1976).
6. Geisler, N. & Weber, K. *Biochemistry* **16**, 938-943 (1977).
7. Ogata, R. T. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5851-5854 (1978).
8. Guinier, A. & Fournet, G. in *Small-angle Scattering of X Rays* (Wiley, New York, 1955).
9. Jacrot, B. *Rep. Prog. Phys.* **39**, 911-953 (1976).
10. Farabaugh, P. L. *Nature* **274**, 765-769 (1978).

11. Serdyuk, I. N., Zaccari, G. & Spirin, A. S. *FEBS Lett.* **94**, 349–352 (1978).
12. Satre, M. & Zaccari, G. *FEBS Lett.* **102**, 244–248 (1979).
13. Ramstein, J., Charlier, M., Maurizot, J. C., Szabo, A. G. & Hélène, C. *Biochem. biophys. Res. Commun.* **88**, 124–129 (1979).
14. Platt, T., Files, J. G. & Weber, K. *J. biol. Chem.* **248**, 110–121 (1973).
15. Alexander, M. E., Burgum, H. A., Noall, R. A., Shaw, M. D. & Matthews, K. S. *Biochem. biophys. Acta* **493**, 367–379 (1977).
16. Ogata, R. T. & Gilbert, W. *J. molec. Biol.* (in the press).
17. Chandrasekaran, R., Jardetzky, T. S. & Jardetzky, O. *FEBS Lett.* **101**, 11–14 (1979).
18. Buck, F., Rüterjans, H. & Beyreuther, K. *FEBS Lett.* **96**, 335–338 (1978).
19. Wade-Jardetzky, N. *et al. J. molec. Biol.* **128**, 259–264 (1979).
20. Steitz, T. A., Richmond, T. J., Wise, D. & Engelman, D. *Proc. natn. Acad. Sci. U.S.A.* **71**, 593–597 (1974).
21. Ohshima, Y., Horiuchi, T. & Yanagida, M. *J. molec. Biol.* **91**, 515–519 (1975).
22. Geisler, N. & Weber, K. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3103–3106 (1976).
23. Kania, J. & Brown T. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3529–3533 (1976).
24. Müller-Hill, B., Beyreuther, K. T. & Gilbert, W. *Meth. Enzym.* **21** part D, 483–487 (1974).

The caries resistance of human teeth is determined by the spatial arrangement of hydroxyapatite microcrystals in the enamel

Gregor Cevc*†, Pavle Cevc†, Milan Schara† & Uroš Skalerič†

* Institute for Biophysics, University E. Kardelj, Ljubljana, Lipičeva 2, 61000 Ljubljana, Yugoslavia

† Institute J. Stefan, University E. Kardelj, Ljubljana, Jamova 39, 61000 Ljubljana, Yugoslavia

‡ Dental Clinic, Medical School, University E. Kardelj, Hrvatski trg 6, 61000 Ljubljana, Yugoslavia

Teeth from different people differ in their susceptibility to caries. The beginning and the development of tooth decay depend on the presence of bacterial plaque¹ and on the structure of the tooth enamel². But it has not been established whether the organic or the inorganic component of the enamel (which is comprised principally of hydroxyapatite microcrystals organized in a firm matrix), is responsible for most of the resistance of teeth to caries. We report here that the electron spin resonance (ESR) lines of the CO_3^{3-} defect of samples of enamel prepared from caries-resistant teeth differ significantly from the corresponding spectra of the caries-sensitive samples. We have traced these differences to the hydroxyapatite microcrystalline alignment in the tooth enamel, and found that this alignment is specific to individuals and is therefore probably nutritionally or genetically determined.

We used samples of enamel $2 \times 2 \times 1$ mm, cut from the human maxillary incisors of adults of different sex and age. The outer, square face was taken from the buccal tooth surface (insert in Fig. 1a). To facilitate comparison, several deciduous teeth were also measured and analysed. We had two groups of adult teeth: those removed (because of periodontal disease) from people with above average resistance to caries (25 samples) and those considered to be caries-sensitive because they were grossly decayed (26 samples). In the latter case, care was taken to prepare ESR samples from areas which were completely unaffected by caries.

All samples were γ -ray irradiated at room temperature to a final dose of 0.5 Mrad to introduce defects into the enamel. The CO_3^{3-} centre was predominant; this arises from electron capture by the CO_3^{2-} radical³, but a weak signal of O^- was also observed. The lines were assigned by comparison of the irradiated enamel ESR spectra with those of synthetic hydroxyapatite. Irradiated samples were placed in sealed quartz-glass tubes for measurement on a Varian E-9 ESR spectrometer. Spectra were recorded with the magnetic field perpendicular to the square face of the sample and thus also perpendicular to the long axis of the tooth. Representative spectra were also recorded from enamel samples powdered in a mortar. The spectrometer was operated in the

X-band range and only the central part of the spectra⁴, corresponding to the CO_3^{3-} centre, was analysed.

The carbonate in the enamel may be contained within the hydroxyapatite lattice or present as a separate phase with counterions^{5,6}. ESR studies of natural carbon-containing apatites have, however, shown that the CO_3^{3-} ESR lines originate from the carbonate radicals CO_3^{2-} which replace PO_4^{3-} in the apatite lattice and are therefore firmly bound to it³. For tooth samples, we compared the angular dependence of the ESR lines of the CO_3^{3-} and O^- defects (the orientation of the latter with respect to the crystal axes being well established from other studies of synthetic hydroxyapatites⁵⁻⁷ and thereby obtained further evidence that the ESR-observable CO_3^{3-} indeed reflect the order within the hydroxyapatite matrix⁴.

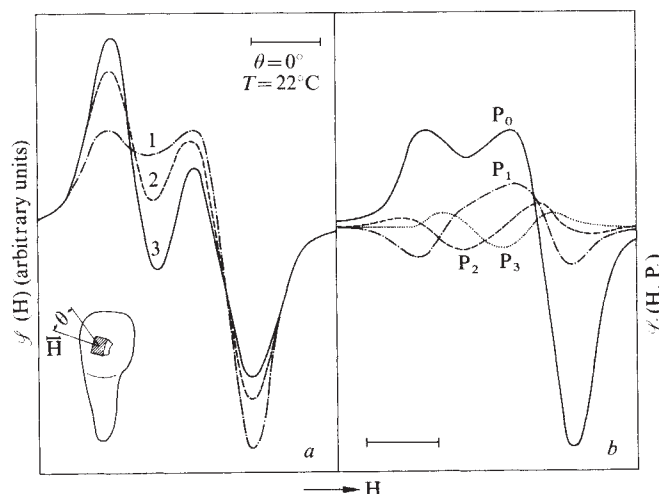


Fig. 1 Measured and simulated CO_3^{3-} centre ESR lines of γ -ray irradiated adult human incisor enamel. a, The experimental spectra of representative powdered (1), caries-sensitive (2) and caries-resistant (3) teeth; b, the first four calculated ESR spectra $P_i(H)$, $i = 1, 2, 3, 4$, which correspond to the angular distribution function of the CO_3^{3-} centre preferential axes given by the lowest four Legendre polynomials. Insert shows the site on the tooth buccal surface from where the samples were cut and their orientation in the magnetic field. Scale bars represent 10 G. $\theta = 0^\circ$; temperature = 22°C .

The ESR spectra of oriented, cut samples differ from spectra of powdered samples (Fig. 1a). As the ESR lines, which arise from the paramagnetic centres, mirror internal order within the sample, it is possible to deduce the orientational distribution function of these centres in the tissue studied. This function is correlated with the probability that the molecular axes of the observed centres are aligned preferentially in a certain direction with respect to the external magnetic field, corresponding to the state of the internal order of the system.

To understand and simulate the ESR spectrum, the energy states of the observed spin, which can be calculated from the values of the anisotropic hyperfine coupling tensor, as well as the lineshape parameters, must be known. In the procedure that we have developed⁸ the ESR spectra of powdered enamel have first been simulated. A linear combination of the gaussian and lorentzian convolution functions⁹ has been used. The optimal parameters have been determined in the standard way by minimizing the squared differences between the measured and the calculated spectra. Then, the ESR spectra of the corresponding partially ordered specimen have been simulated: the ESR line-determining parameters used were those obtained from the simulation of the powdered enamel spectra, and the Legendre polynomials¹⁰ $P_i(\theta)$, $i = 1, 2, \dots, 12$ have been

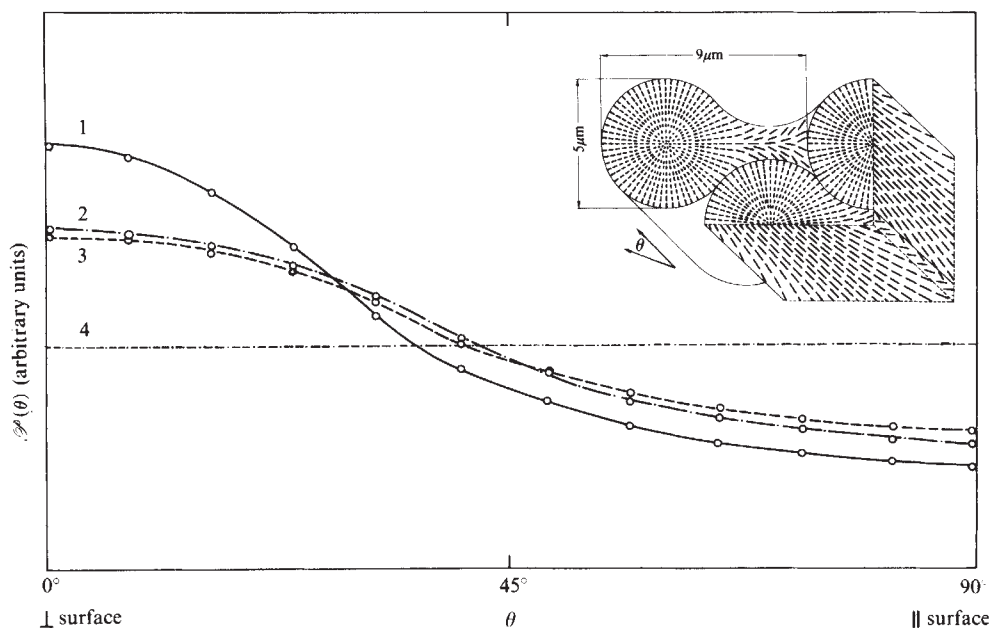


Fig. 2 The angular distribution function of the hydroxyapatite microcrystals within human incisor enamel. Their alignment is defined with respect to the normal to the tooth surface. The points have been obtained by summing the values of Legendre polynomials at respective angles and weighted by the factors α_i , defined in equation (1), and determined from the simulation of the tooth enamel ESR spectra. The curves correspond to the caries-resistant (1), caries-sensitive (2), milk (3) and powdered (4) human teeth. The insert presents schematically the arrangement of the hydroxyapatite microcrystals in lockhole-shaped tooth enamel prisms which are the basic units of the enamel¹³.

assumed to represent the angular distribution function of the CO_3^{3-} centres — and therefore of the hydroxyapatite microcrystals — in the sample. The angle ϑ was defined with respect to the external magnetic field, that is, normal to the enamel surface (insert in Fig. 1a). We have calculated a set of the spectra $\mathcal{S}_i(H) = \mathcal{S}_i(H, P_i(\vartheta))$, $i = 1, 2, \dots, 12$, (Fig. 1b), which have been normalized properly¹¹ as a basis for the expansion of the experimental spectrum of the tooth enamel $\mathcal{P}(H)$ into the series

$$\mathcal{P}(H) = \sum_{i=1}^{12} \alpha_i \mathcal{S}_i(H) \quad (1)$$

The factors α_i in the above sum therefore define the relative contributions of the hypothetical samples of known microcrystalline alignment, given by $P_i(\vartheta)$, in the experimental ESR spectrum. This macroscopic sample spectrum is actually a result of the overlap of the particular ESR lines corresponding to the single hydroxyapatite microcrystals. The factors α_i therefore also reveal the relative amounts of the Legendre polynomials $P_i(\vartheta)$ used for obtaining $\mathcal{S}_i(H)$ in the actual distribution function of these hydroxyapatite microcrystals in the bulk sample $\mathcal{P}(\vartheta)$

$$\mathcal{P}(\vartheta) = \sum_{i=1}^{12} \alpha_i P_i(\vartheta) \quad (2)$$

We conclude that the differences in the CO_3^{3-} centre ESR lines of the teeth studied (Fig. 1a), which correlate significantly to their resistance to tooth decay⁴, are a consequence of a higher degree of alignment of the hydroxyapatite microcrystals in the enamel of the caries-resistant teeth as compared to caries-sensitive teeth.

The electron microscope studies of etched enamel have shown that the hydroxyapatite in the human enamel is ordered in well defined patterns^{12,13}. Around the areas of well aligned microcrystals there are zones of lesser microcrystalline order. As yet, such studies have only yielded localized information about the tooth surface whereas our approach permits conclusions to be drawn about the order averaged over the whole sample, and can thus be compared with the neutron diffraction studies of apatite orientation in bones¹⁴. In this study we have found that the hydroxyapatite order, as revealed by the angular distribution function of microcrystals, extends throughout the outer layer of the enamel (Fig. 2). In this layer, which acts as the main tooth-protecting barrier, the enamel seems to be organized in

such a way that the long axes of microcrystals are on average perpendicular to the tooth surface, an arrangement which confers optimal resistance to the detrimental influences of the oral environment (insert in Fig. 2). The more randomly oriented microcrystals (which contribute to the flat portion of the distribution function, $\vartheta = 45^\circ$ – 90°) would imply little resistance of the corresponding regions to caries. However, these areas constitute the inner enamel parts and are therefore less significant for the overall sensitivity of the tooth to caries: the ESR spectra of the consecutive layers of tooth enamel show a decrease in alignment of microcrystals with distance from the tooth surface⁴. It may also be significant that milk teeth have a poorly organized hydroxyapatite matrix and in this respect resemble caries-sensitive teeth. The limited lifetime of milk teeth suggests that they need not be very resistant to caries (Fig. 2). In caries-resistant teeth, however, a larger proportion of the microcrystals is well oriented. In the angular distribution function this better alignment of hydroxyapatite in the outer enamel region is partially screened by the large background of more randomly oriented inner parts of bulky samples.

The alignment of the microcrystals in the hydroxyapatite matrix of human tooth enamel of different incisors of the same subject are alike⁴. We therefore conclude that the degree of hydroxyapatite microcrystal alignment is an individual property which may, for each individual, determine resistance to tooth decay. However, it remains to be seen whether nutritional or genetic factors, or both, are responsible.

Received 22 January; accepted 21 May 1980.

- Kleinberg, J. in *Oral Microbiology* 3rd ed. (ed. Nolte, W. A.) 515–534 (Mosby, St Louis, 1977).
- Frank, R. M. in *Caries Resistant Teeth* (eds Wolstenholme, G. E. & O'Connor, M.) 169–184 (Churchill, London, 1965).
- Gilinskaya, L. G., Shcherbakova, M. Ya. & Zanin, Yu. N. *Kristallografiya* **15**, 1016–1019 (1971).
- Cevc, P., Schara, M., Ravník, Č. & Skalerič, U. *J. dent. Res.* **55**, 691–695 (1976); *Radiat. Res.* **51**, 581–589 (1972).
- Elliot, J. C. *Calcium Tissue Res.* **3**, 293–307 (1969).
- Elliot, J. C., Mackie, P. E. & Young, R. A. *Science* **180**, 1055–1057 (1973).
- Mengeot, M., Bartram, R. H. & Gilliam, O. R. *Phys. Rev.* **B11**, 4110–4124 (1975).
- Cevc, G., Cevc, P., Schara, M., Ravník, Č. & Skalerič, U. *Period. Biol.* **78**, 178–179 (1976).
- Abragam, A. & Bleaney, B. in *Electron Paramagnetic Resonance of Transition Ions*, 54–59 (Clarendon, Oxford, 1970).
- Abramovitz, M. & Stegun, J. A. in *Handbook of Mathematical Functions*, 473 (Dover, New York, 1971).
- Hentschel, R., Schlitter, J., Sillescu, H. & Spiess, H. W. *J. chem. Phys.* **68**, 56–66 (1978).
- Hamilton, W. J. Jr, Judd, G. & Ansell, G. S. *J. dent. Res.* **51**, 1407–1420 (1972).
- Scott, D. D., Simmelink, J. W. & Nygaard, V. *J. dent. Res.* **53**, 165–178 (1974).
- Bacon, P. J. & Griffith, R. K. *J. appl. Crystallogr.* **12**, 99–103 (1979).

BOOK REVIEWS

Trials of Jewry

Vera Rich

OVER the past ten years, the problems facing Soviet Jews in general and Soviet Jewish scientists in particular have frequently been raised in both the general and the scientific press. To a large extent, however, this has been a matter of Jews wishing to emigrate to Israel: Jews, it may be argued, who have chosen to opt out of the Soviet system. What is less often realized is that in one branch of Soviet life there appears to be active discrimination against Jews, however little they themselves feel aware of their Jewish identity — namely in mathematics.

It Seems I Am a Jew, originally a *samizdat* (privately circulated) essay and now published in English translation, is basically a record of one man's experiences in trying to make for himself a career in mathematics while handicapped by the implications of the word 'Jew' in his internal passport, under Point 5, 'Nationality'. For, from the point of view of the Soviet administration, Jewishness relates not to religion, not to self-identification with an ancestral cultural tradition, but simply to ethnic descent. Thus in the case of some anti-Jewish prejudice or campaign, as with the notorious 'Doctors' plot' in the last days of Stalin, any person of Jewish descent, no matter how *déraciné*, is liable to fall victim.

Such prejudices or campaigns are, of course, totally contrary to the Soviet constitution. The Soviet press constantly reiterates that there is no 'anti-Semitism' in the Soviet Union and that Soviet policy is opposed only to 'Zionism'. Nevertheless, as Freiman's essay indicates, there is growing evidence that discrimination against Jews, particularly in the academic professions, and especially in mathematics, does exist. Promotions over one's head, unreasonable delays in the acceptance of a doctoral thesis, refusal (or at least non-provision) of a passport for foreign travel — these are phenomena all too common in the life of the Soviet Jewish intellectual. (Ironically, such incidents have sometimes driven a hitherto uncommitted Jew right into the arms of the emigration lobby. Aleksandr Ioffe, whose mathematical work recently formed the basis of a two-day seminar in London, had shown no signs of wanting to emigrate to Israel — until he was refused a passport to take a temporary post as 'visiting lecturer' in Paris.)

To such a situation, which could so easily become fraught and over-emotional, Freiman brings a rare, if black, humour. The 'documentary poem' of his attempts in

It Seems I Am a Jew. By Grigori Freiman. Pp.98. (Southern Illinois University Press: Carbondale, Illinois/Feffer and Simons: London, 1980.) \$9.95, £5.45.



Grigori Freiman

the early 1950s to enrol as an external graduate student of mathematics, while working at the Kazan Institute of Aviation, is a masterly analysis of 'Catch-22' bureaucracy. Yet since that time, he says, the situation has significantly deteriorated, with admission of Jews to the more prestigious mathematical departments and institutes being phased out almost completely since 1970. The universities of Moscow, Leningrad and Novosibirsk, the main stepping-stones to a career in creative mathematics, have, says Freiman, virtually ceased to admit Jews. He describes a special selection process by which Jews and certain other 'undesirables', such as Armenians, are concentrated for the entrance examination into special groups, which are then set far more difficult problems than the racially more acceptable candidates. Examples of these 'Jewish problems' are given in the text; further examples, from the examination of 1978 for the Moscow University Faculty of Mechanics and Mathematics, are cited in an appendix. Although there has been

previous mention of such 'Jewish problems' in print, this is the first time that a sufficient number have been cited to make it possible to assess them.

Judging from these, for 'special selection' the examiners seem to favour number theory and classical geometry, subjects in which intuition plays a particularly significant role. One or two of the questions, incidentally, notably the solution in integers of the equation $x^y = y^x$, can be solved intuitively and virtually instantaneously — in the context of pseudo-mathematical party-games. If, however, a rigorous mathematical proof must be produced — orally, within 15 minutes, under the pressure of examination conditions, by a nervous 19 year old — the situation takes on a rather different aspect. The standard Soviet explanation of such questions, as Freiman points out, is that the examiners cannot admit everyone, but are on the look out for excellence. In that case, he asks, why do Jewish students who have come out in the top bracket of the All-Union Mathematical 'Olympiads' do so badly in the entrance examinations? (And why, one may add, are such supplementary tests of excellence not administered to all of the candidates?)

Yet the entrance examination is only the first of the problems to be faced in the academic careers of Soviet Jews. Not only are they refused admission to the best universities, but at every step their advancement is blocked. The thesis is turned down, not because the aspirant is called Goldberg or Finkelstein, but because it was not quite up to standard. The promotion bypassed Lazar Izaakovich in favour of Ivan Petrovich not on grounds of nationality, but simply because the ethnically Russian candidate happens to be the better, more qualified and more capable applicant. Such 'explanations' are notoriously difficult to refute — all the more so since, according to Freiman, the examiners, expert committees and appointments' boards all have to play a double game; knowing that they must reject a Jewish candidate, they have to produce an academic alibi. "The mathematics is mechanical . . ."; "He is a brilliant mathematician but superficial"; "He has about a hundred works to his name. He must be chasing after recognition". Such gems, taken, says Freiman, from the official stenographic reports of the various assessing bodies, show only too clearly the cynical nature of their verdicts.

Freiman's essay is essentially a personal statement, founded on his own experience

Since this review was written, Professor Freiman has been dismissed from his post at Kalinin State University.

and on data which he has been able to garner from acquaintances. He does not attempt to explain or find psychological or political motivation behind the drive to render Soviet mathematics *judenfrei*. He is content simply to state the facts as he sees them, and to cite, in an impassioned *J'accuse* addressed to the President of the Soviet Academy of Sciences, the names of half-a-dozen leading Soviet mathematicians, the Expert Committee for Mathematics and Mechanics of the Higher Attestation Committee (the body responsible for higher degrees) and the Steklov Institute of Mathematics of the Soviet Academy of Sciences. These few put prestigious persons and bodies, he says, have made Soviet

mathematics "a cesspool of scientific unprincipledness".

Granted the conditions described in Freiman's essay, and the resolute official denials that anything in the nature of anti-Semitism can exist in the Soviet Union, it would seem virtually impossible for any impartial third party to assess the validity of Freiman's claims. A certain weight should undoubtedly be given to the fact that at the time of writing the essay Freiman himself was neither a refusenik nor an *émigré*. (Since then, he has filed an application to emigrate to Israel.) Weight, too, should be given to the style of the essay, remarkably free from the introspective self-pity that mars so many *samizdat* biographies of

repression. One piece of evidence mentioned by Freiman is, however, open to checking by the Western reader — the virtual disappearance over the past few years of Jewish surnames from the contributors to the *Matematicheskii Sbornik* and other leading Soviet mathematical journals. Thanks to the various cover-to-cover translation bureaux, most of these journals are available, in their entirety, in English translations. Those readers who find Freiman's revelations difficult to accept can count for themselves. □

Vera Rich is a journalist who specializes in East European science affairs.

Publicity for exotic atoms

C.J. Batty

Exotic Atoms '79: Fundamental Interactions and Structure of Matter. Edited by Kenneth Crowe, Jean Duclos, Giovanni Fiorentini and Gabriele Torelli. Pp.403. (Plenum: New York and London, 1980.) \$45, £28.35.

AN EXOTIC atom is an atom in which one of the electrons has been replaced by a negatively charged heavy particle such as a muon (μ), π or K-meson, antiproton ($\bar{p}$) or by a Σ -hyperon. It can be formed by stopping a beam of the appropriate particles in a suitable nuclear target. Once the heavy particle has been slowed down to an energy of order 1 keV it can be captured into a Bohr orbit about the nucleus of high principal quantum number n . It will then cascade down through its own sequence of atomic levels, first by Auger transitions with the emission of an electron and later by the emission of X-rays. Following each of these transitions it comes closer to the nucleus until at some stage nuclear capture takes place.

Studies of exotic atoms cover a whole range of disciplines from physical chemistry through atomic and solid-state physics to nuclear and elementary particle physics. This diversity of interests and applications is reflected in the 22 articles, of varying length, which appear in *Exotic Atoms '79*.

The exact mechanism by which the heavy particle is captured to form the exotic atom is not fully understood but the present volume contains an interesting review of this topic by Schneuwly under the heading of "Mesic Chemistry". Four other shorter articles deal with more specific aspects of capture by hydrogen and the properties of muonic-hydrogen atoms. A particularly fascinating review by J. Rafelski from CERN deals with the properties of mesomolecules of hydrogen in which the electron in the hydrogen molecule H_2^+ has

been replaced by a muon. In the case of the heavier hydrogen isotopes spontaneous fusion can occur, a process first observed for the system μ^-pd at Berkeley in 1956, and there is now considerable interest in the properties of dd and dt mesomolecules.

Rather special examples of exotic atoms are muonium, the system μ^+e^- , and positronium, the system e^-e^+ . The properties of muonium and positronium are reviewed by V. Hughes of Yale University, who has worked for many years in this field, and here one is particularly impressed by the elegance of the experiments and the precision of the results. Extensions of the exotic atom picture to consider bound states of quarks to form hadrons are discussed in three theoretical articles. Other topics reviewed include muon capture rates and parity non-conservation in atoms and molecules.

The whole field of muon spin rotation (μSR), so named because of its close analogy with nuclear magnetic resonance

(NMR) and electron spin resonance (ESR), has expanded dramatically over the last few years and is now a valuable tool in the study of condensed matter as well as being a topic of study in its own right. It is therefore not surprising that almost half of the book is devoted to this subject. Topics covered include muon diffusion and trapping in solids, aspects of positive muons as impurities in metals, studies of μSR in semiconductors and in ferromagnetic metals, and some features of muonium chemistry.

The book is certainly a valuable contribution to publicizing the existence of a fascinating area of cross-disciplinary research, and will provide an excellent introduction to those meeting the subject for the first time and a valuable reference work for the more expert. □

C.J. Batty is on the staff of the Rutherford and Appleton Laboratories at Chilton, Oxfordshire, UK.

Neurophysiological methods in entomology

Melody Siegler

Insect Neurophysiological Techniques. By T.A. Miller. Pp.308. (Springer: Heidelberg and New York, 1979.) DM54, \$29.70.

IN *Insect Neurophysiological Techniques*, T.A. Miller has compiled details of techniques taken from the published works of many authors, and from his own visits to and correspondence with several laboratories. The text is divided into four parts. I, "Materials", describes electrophysiological apparatus and accessories; II, "Actographs", reports the methods used over the past 25 years to measure insect activity; III, "Free-Moving and Tethered Preparations", describes the

procedures selected researchers have used to record from nerve and muscle in cockroaches, locusts, crickets and flies; IV, "Organ and Tissue Preparations", emphasizes heart, gut and salivary gland preparations, methods of recording from cockroach ventral nerve cord, and nerve-muscle preparations of locust and cockroach leg. There is a reference section, containing selected literature prior to 1977, an appendix listing suppliers and an index. The book is the first in a series to be concerned with problems of experimental entomology.

Ideally, a book that emphasizes methodology should be accurate, clear, concise and well-organized, and be designed to make its contents readily accessible to the reader. *Insect Neurophysiological Techniques* falls far short of these ideals. The text is wordy, and almost awash with minutiae. Often it strays from the topic at hand, and large sections are misplaced or mistitled. There is a wealth of detail, but it is often repetitive or

with no clear objective. I wish the author had exerted more of an organizing hand, carefully and firmly integrating the information from his diverse sources. Throughout, the level of presentation is uneven. For some topics, the writing is addressed to those readers with little or no laboratory experience; for others there is little introduction and much jargon. Too often minor points are given misleading emphasis, whilst crucial details are

incorrect or omitted altogether. The dedicated and cautious reader will doubtless be able to extract some useful information but, in general, the book is not a particularly helpful introduction to the considerable advantages of insects for basic or applied research. □

Melody Siegler is a Research Fellow in the Department of Zoology, University of Cambridge, UK.

Insight into microbial development

Ian Dawes

Developmental Biology of Prokaryotes. Edited by J.H. Parish. Pp.297. (Blackwell Scientific: Oxford, 1980.) £19, \$48.50.

THERE are few texts available that cover the many aspects of microbial development and those that have been around for rather too many years to be of much use to students or teachers. There is, therefore, a need for a suitable text at a fairly advanced level and this book is an attempt to fill the gap.

It tries to do this by way of a brief introduction by the editor to prokaryotic diversity and unity, and to gene expression and its regulation. This is followed by ten chapters covering nine groups of organisms, each chapter written by an expert in his particular area. These include the cell cycle in *Escherichia coli*; bacterial sporulation (with two chapters in deference to the volume of research that has been carried out in this area); sphere-rod transitions in *Arthrobacter*; streptomycete differentiation; *Bdellovibrio* growth; prosthecate bacteria; cyanobacterial differentiation (filamentous and pleurocapsalean); and myxobacteria. A brief concluding chapter rounds off the book.

Foremost among the difficulties in trying to provide a text by this type of approach is the tendency to produce a collection of individual review articles, each a catalogue of the hopes, successes and failures of research in one particular area, at the expense of formulating a fundamental analysis of the subject. Is this not integrating differentiation? This book is no exception in that the two chapters that could provide an overview of the subject are too condensed to be of much value; much of the introduction is best dispensed with and the reader should consult, where necessary, the appropriate biochemistry bible of choice. The majority of material has, however, been chosen judiciously with a balanced coverage of the main bacterial development systems, although in future editions several chapters on bacteriophage development could be added to illustrate

relevant control and assembly systems that have been worked out in considerable detail.

Several of the chapters are excellent. That on the cell cycle is stimulating, both for its breadth of coverage and for the critical discussion of possible models to account for the mechanisms controlling DNA replication and cell division. The two chapters on bacterial spores and sporulation also provide interesting reading despite the plethora of reviews already available on the subject. Inevitably, that concerned with the molecular biology of sporulation is beginning to date, as gene cloning techniques are rapidly exploited. Nonetheless, the chapter provides a useful summary of the main work that has been done on the control of spore formation. The authors of the remaining chapters have faced the difficulty (recognized by Parish) that "there is a lot of work to do on other groups of 'differentiating prokaryotes' to bring the state of knowledge up to the level of *B. subtilis*". In most cases, they have succeeded in avoiding the temptation to compare progress with that in *B. subtilis* by stressing the aspects of development that are unique to their system. There are occasional lapses, none more revealing than the several discussions of the lack of evidence for RNA polymerase changes during differentiation in *Caulobacter* and *Streptomyces*. One or two chapters lack freshness since the authors have little to add to recent reviews on the same subjects. Also noticeable in places is the inverse correlation between the number of pictures and the number of facts available for presentation. Also, at times, and to the greatest extent in the chapter on *Arthrobacter*, there is a tendency to catalogue biochemical studies, without regard to their relevance or completeness; this makes for less interesting reading.

These are, however, minor criticisms of a book that is valuable for the insight it gives into microbial development processes. Whilst not a text in the sense that it tries to unify the field of prokaryotic development, it provides excellent coverage of the main bacterial systems and should prove very useful to those who study, or teach, microbial development. □

Ian Dawes is a Lecturer in the Department of Microbiology, University of Edinburgh, UK.

A holistic approach to vision

Talbot H. Waterman

The Ecology of Vision. By J.N. Lythgoe. Pp.244. (Clarendon/Oxford University Press: Oxford, 1979.) £22, \$49.

IN THE ecology of vision John Lythgoe has chosen a fascinating topic for a book. Obviously radiant energy mainly from the Sun has had far reaching effects on the nature of the Earth and its fitness for life. More directly, sunlight supplies practically all of the energy needed by organisms to survive, grow and evolve. In the complementary field of biological information and control natural light also functions everywhere in daily and seasonal programming as well as in the spatial orientation of both plants and animals.

Limiting one's attention to the photobiology of animals with well-developed functional eyes as implied by 'vision' still leaves an enormously challenging area to be dealt with. Lying as it does at the intersection of comparative visual physiology and ecology it has long been the major area of my own research and teaching. Indeed I would have liked to write a book about this subject myself. Hence this appraisal of Lythgoe's work, while it may betray a trace of sour grapes, cannot fairly be attributed to disinterest.

However, coping well with the taxing demands of so broad and varied a subject is a formidable goal. Starting with solar irradiation, the path leads insistently through meteorological as well as submarine optics to the biophysics and biochemistry of visual pigments, the functional organization of eyes, thence to visually mediated perception, orientation, navigation, migration and behaviour.

Obviously most of these component topics have often been reviewed in their own right, sometimes with a high degree of competence. Even in the holistic sense there are several symposium volumes that have essayed to survey the substantial area where natural radiation, ecology and sensory physiology come together.

Thus the author of a book on this subject not only has a big topic to deal with but he also has to encompass the previous contributions of numerous other scientists. Hence the questions: "What should be said that hasn't already been well said?" "What is the real *point d'appui* of the present approach?" and finally, "To whom is the work addressed?", are not trivial ones easily answered.

An unusually versatile author working resolutely with consummate skill might solve most of these problems if overall strong matching support was provided by his publisher. Yet weakness or failure in fulfilling one or more of the book's numerous desiderata will at best result in a 'nice try' rather than a new classic. In my

opinion the present volume, while certainly stimulating, is not unflawed and has proved to be rather disappointing.

On the positive side *The Ecology of Vision* has, in addition to its subject, other strong points in its favour. Thus, the overall organization is fine. In outline there are two introductory chapters (80 pp.) on the photic environment and vision respectively. A middle section (89 pp.), in four chapters, details the visual consequences of (1) low ambient light intensities (night, deep water, caves), (2) high turbidity of the medium (which often restricts aerial vision and always puts a severe limit on underwater visibility), (3) the interaction of directional illumination and preferred lines of sight, and (4) spectral bias of non-white illumination (e.g. sky light, mesopelagic irradiance and reflected light from vegetation). The final chapter (35 pp.) surveys the organism-environment interaction as it relates to camouflage, conspicuousness and recognition.

Thus the text overall is brief, and

therefore appropriate to a broad readership which might not persist through a multi-authored encyclopaedic treatise. Yet Lythgoe has managed to include so varied a coverage that even the specialist is bound to find some fascinating correlations previously unappreciated. These fresh insights, however, may seem to lie around the periphery rather than near the centre of a given reader's area of expertise. For instance two topics of primary interest to me, compound eyes and polarization sensitivity, seemed among the less effective aspects of the book.

A further and notable positive feature of *The Ecology of Vision* is its wealth of illustrative material. This eye-catching potpourri of line drawings and colour as well as black-and-white half-tones is attractive both to the general reader and to the specialist, ranging from portrait sketches of many of the different animals cited in the text to excellent illustrations chosen from relevant research publications including Lythgoe's own. While the idea of illustrating many of the organisms referred

to in this comparative study is welcome, its impact is blunted if adequate biological and aesthetic standards are not maintained. Rather rudimentary drawings of a six-legged spider, a crayfish with eight legs and a bifurcate rostrum, as well as a gecko apparently fitted with dentures, are extreme examples in point.

The bibliography, of 340 items, is reasonably up to date; however one disturbing aspect of the references is an almost total failure of the author to guide the reader to the most important sources of further original information. Thus few primary publications of innovative research workers are presented in any sort of historical or evaluative perspective. Lythgoe is surely not alone in this tendency to become lax or even irresponsible in identifying the seminal contributions to a field of science. Further, a number of experimental biologists whose contributions would loom large in some readers' view of the ecology of vision are not even mentioned: Karl von Frisch, Erich von Holst, H.K. Hartline and Werner Reichardt, for example.

This brings up another feature of the book which is particularly evident in the bibliography, namely its marked skewness towards the vertebrates. Sixty percent of the citations refer to this group and of these 57% in turn are references to fishes. This bias reflects the author's own research interest which has mainly been concerned with underwater visibility and fish vision.

Even with regard to fishes, major seemingly relevant topics are ignored. For instance the dorsal light reflex, sun compass behaviour, diurnal vertical migration, light-cued rhythms and photoperiodic influences, as well as the whole field of chromatophores and their control, are all omitted. Obviously the coverage of insects, spiders, crustaceans and cephalopods, all of which have highly developed vision, is even less representative of the great amount of research which has been done on their visual ecology.

One pervasive feature of the text suggests that the author may have surrendered his manuscript to the publisher before its due term. This is a certain lack of polish and style in the writing. In some sections repetitive and awkwardly coordinated passages are common, while the style at its least elegant may be epitomized by the disappointing introduction to Chapter 1 and the rather limp coda at the very end.

How can such a good idea for a book have materialized as unevenly as this? John Lythgoe provided an important and interesting topic generally well organized, well illustrated and well documented. But in spite of this the book fails to live up to its promise. The flaws are too many for it to be regarded as a prospective classic, though it is certainly a 'nice try'.

Brain structures and behaviour

Richard Lynn

Brain and Personality. By Graham E. Powell. Pp. 113. (Saxon House: Farnborough, UK, 1979.) £10.50.

IN THIS volume Dr Powell reviews the current state of theory and knowledge of the relation between brain structures and personality. The methodology in this field consists primarily of making lesions in different parts of the brain and observing subsequent changes in the personality of the subject. The first substantial advance was the documentation in the nineteenth century of the case of Phineas Gage, a construction worker whose frontal lobes were severely damaged in an accident and who subsequently became more happy-go-lucky and less socially responsible. In later years this type of effect has frequently been reported following frontal lobe injury.

Dr Powell argues that to make progress in the field it is necessary to have a model of the major personality traits or dimensions. We assume that different brain structures serve particular personality traits. The brain can then be mapped by observing the effects of lesions in different parts of the brain on various personality traits. Dr Powell begins with Professor Eysenck's traits of neuroticism and extraversion and his theory that these are mediated by the thalamic and brain stem reticular formations. This formulation has encountered certain difficulties. An attempt to meet them has been made by J. Gray (*The Psychology of Fear and Stress*:

1971) who has proposed that introverts are more sensitive to punishment and that this sensitivity is mediated by the orbital frontal cortex, the medial septal area and the hippocampus. Here again, however, the new theory has met difficulties. In particular, it is still not clear whether damage to this system really does make people less introverted.

The author turns next to the more empirical studies. These consist of attempts to relate lesions of various brain structures to personality change, without any model of brain function. The chief results here are the reduction of aggressiveness following damage to the amygdala, the reduction of anxiety following damage to the cingulate gyrus and the reduction of sexual drive following damage to the hypothalamus. These findings have been used in the USA to reduce criminal aggressive and sexual behaviour. This is a useful source of knowledge of the brain mechanisms underlying these behaviours, although not one of which the author approves.

Looking at work on these problems in the century or so that has elapsed since the Phineas Gage case, one is left with a certain feeling of disappointment about the slowness of the progress that has been made. Perhaps the use of questionnaires such as Cattell's (*Personality and Learning Theory*: 1979), which measure narrower personality traits, would yield more decisive results. In the meantime, Dr Powell's book will be a useful starting point for those attempting to come to grips with the still largely unsolved problems in this field.

Richard Lynn is Professor of Psychology at the University of Ulster, UK.

Talbot H. Waterman is Professor of Biology at Yale University, New Haven, Connecticut.